```
   1 CGAGTGCGCT GCTGAGTCCT GTAGATAAAG CCGCCAACCC CGGGGACTGG
  51 TGTCTCCTGA GTGACCGTGC AGCCGTGGGC GCCATAGAAA GCAGAGAAGG
 101 CAGTGAACTT CGACCACTTC CAGATCCTTC GGGCCATTGG GAAGGGCAGC
 151 TTTGGCAAGG TGTGCATTGT GCAGAAGCGG GACACGGAGA AGATGTACGC
 201 CATGAAGTAC ATGAACAAGC AGCAGTGCAT CGAGCGCGAC GAGGTCCGCA
 251 ACGTCTTCCG GGAGCTGGAG ATCCTGCAGG AGATCGAGCA CGTCTTCCTG
 301 GTGAACCTCT GGTACTCCTT CCAGGACGAG GAGGACATGT TCATGGTCGT
 351 GGACCTGCTA CTGGGCGGGG ACCTGCGCTA CCACCTGCAG CAGAACGTGC
 401 AGTTCTCCGA GGACACGGTG AGGCTGTACA TCTGCGAGAT GGCACTGGCT
 451 CTGGACTACC TGCGCGGCCA GCACATCATC CACAGAGATG TCAAGCCTGA
 501 CAACATTCTC CTGGATGAGA GAGGACATGC ACACCTGACC GACTTCAACA
 551 TTGCCACCAT CATCAAGGAC GGGGAGCGGG CGACGGCATT AGCAGGCACC
 601 AAGCCGTACA TGGCTCCGGA GATCTTCCAC TCTTTTGTCA ACGGCGGGAC
 651 CGGCTACTCC TTCGAGGTGG ACTGGTGGTC GGTGGGGGTG ATGGCCTATG
 701 AGCTGCTGCG AGGATGGAGG CCCTATGACA TCCACTCCAG CAACGCCGTG
 751 GAGTCCCTGG TGCAGCTGTT CAGCACCGTG AGCGTCCAGT ATGTCCCCAC
 801 GTGGTCCAAG GAGATGGTGG CCTTGCTGCG GAAGCTCCTC ACTGTGAACC
 851 CCGAGCACCG GCTCTCCAGC CTCCAGGACG TGCAGGCAGC CCCGGCGCTG
 901 GCCGGCGTGC TGTGGGACCA CCTGAGCGAG AAGAGGGTGG AGCCGGGCTT
 951 CGTGCCCAAC AAAGGCCGTC TGCACTGCGA CCCCACCTTT GAGCTGGAGG
1001 AGATGATCCT GGAGTCCAGG CCCCTGCACA AGAAGAAGAA GCGCCTGGCC
1051 AAGAACAAGT CCCGGGACAA CAGCAGGGAC AGCTCCCAGT CCGAGAATGA
1101 CTATCTTCAA GACTGCCTCG ATGCCATCCA GCAAGACTTC GTGATTTTTA
1151 ACAGAGAAAA GCTGAAGAGG AGCCAGGACC TCCCGAGGGA GCCTCTCCCC
1201 GCCCCTGAGT CCAGGGATGC TGCGGAGCCT GTGGAGGACG AGGCGGAACG
1251 CTCCGCCCTG CCCATGTGCG GCCCCATTTG CCCCTCGGCC GGGAGCGGCT
1301 AGGCCGGGAC GCCCGTGGTC CTCACCCCTT GAGCTGCTTT GGAGACTCGG
1351 CTGCCAGAGG GAGGGCCATG GGCCGAGGCC TGGCATTCAC GTTCCCACCC
1401 AGCCTGCTGC GCGGTGCCCA CAGTGCCCCG GACACATTTC ACACCTCAGG
1451 CTCGTGGTGG TGCAGGGGAC AAGAGGCTGT GGGTGCAGGG GACACCTGTG
1501 GAGGGCATTT CCCGTGGGCC CCCGAGACCC GCCTAGATGG AGGAAGCGCT
1551 GCTGGGCGCC CTCTTACCGC TCACGGGGAC CTGGGGGACA GGATGGGACA
1601 GGAGTCTTTG TCCCTGCTCA GCCCGGAGGC TGTGCACGGC CCTCGTCACA
1651 AGGTGACCCT TGCAGCACAG GCCGCGGGTG CCCCAGGCTC GGCTCAGTTC
1701 TTGGAGGTCA AGGGCATGGG TTGGGGTAGT GGGTGGGGAG GTGAATGTTT
1751 TCTAGAGATT CAAACTGCTC CAGCAATTTC TGTATAGTTT TCACCTCTGA
1801 GAATTACAAT GTGAGAACCG CACAAAAAAA AAAAAAAAAA AAAAAAAAAA
1851 AAAAAAAAAA AAAA
```
(SEQ ID NO: 1)

FEATURES:
5'UTR: 1 - 192
Start Codon: 193
Stop Codon: 1300
3'UTR: 1303

Homologous proteins:
Top 10 BLAST Hits:

Sequences producing significant alignments:	Score (bits)	E value
CRA|103000001515936 /altid=gi|10946600 /def=ref|NP_067277.1| hy...	760	0.0
CRA|120000042903164 /altid=gi|13358640 /def=dbj|BAB33045.1| (AB...	733	0.0
CRA|87000000001314 /altid=gi|8923754 /def=ref|NP_060871.1| gene...	556	e-157
CRA|87000000001426 /altid=gi|7161864 /def=emb|CAB76566.1| (AJ25...	554	e-156
CRA|108000024647823 /altid=gi|12730486 /def=ref|XP_003392.2| ge...	423	e-117
CRA|18000005184360 /altid=gi|7505957 /def=pir||T23688 hypotheti...	335	1e-90
CRA|18000005004115 /altid=gi|1730069 /def=sp|P54644|KRAC_DICDI ...	217	4e-55
CRA|18000004912236 /altid=gi|464395 /def=sp|P28178|PK2_DICDI PR...	203	8e-51
CRA|18000004910302 /altid=gi|1170689 /def=sp|P42818|KPK1_ARATH ...	202	1e-50
CRA|89000000197925 /altid=gi|7295638 /def=gb|AAF50945.1| (AE003...	201	2e-50

FIGURE 1A

<u>EST:</u>

```
                                                          Score      E
Sequences producing significant alignments:             (bits)   Value
gi|13032240 /dataset=dbest /taxon=960...                  1348   0.0
gi|6588496 /dataset=dbest /taxon=9606 ...                 1021   0.0
gi|883123 /dataset=dbest /taxon=9606 /...                  702   0.0
gi|946492 /dataset=dbest /taxon=9606 /...                  236   1e-59
```

EXPRESSION INFORMATION FOR MODULATORY USE:
gi|13032240 prostate
gi|6588496 /lung
gi|883123 /whole brain
gi|946492 / Adult brain

<u>Tissue Expression</u>
whole brain

FIGURE 1B

```
  1 MYAMKYMNKQ QCIERDEVRN VFRELEILQE IEHVFLVNLW YSFQDEEDMF
 51 MVVDLLLGGD LRYHLQQNVQ FSEDTVRLYI CEMALALDYL RGQHIIHRDV
101 KPDNILLDER GHAHLTDFNI ATIIKDGERA TALAGTKPYM APEIFHSFVN
151 GGTGYSFEVD WWSVGVMAYE LLRGWRPYDI HSSNAVESLV QLFSTVSVQY
201 VPTWSKEMVA LLRKLLTVNP EHRLSSLQDV QAAPALAGVL WDHLSEKRVE
251 PGFVPNKGRL HCDPTFELEE MILESRPLHK KKKRLAKNKS RDNSRDSSQS
301 ENDYLQDCLD AIQQDFVIFN REKLKRSQDL PREPLPAPES RDAAEPVEDE
351 AERSALPMCG PICPSAGSG
    (SEQ ID NO: 2)
```

FEATURES:
Functional domains and key regions:
[1] PDOC00001 PS00001 ASN_GLYCOSYLATION
N-glycosylation site

        288-291 NKSR
----
[2] PDOC00005 PS00005 PKC_PHOSPHO_SITE
Protein kinase C phosphorylation site

Number of matches: 2
        1   75-77 TVR
        2   245-247 SEK
----
[3] PDOC00006 PS00006 CK2_PHOSPHO_SITE
Casein kinase II phosphorylation site

Number of matches: 4
        1   42-45 SFQD
        2   226-229 SLQD
        3   298-301 SQSE
        4   300-303 SEND
----
[4] PDOC00008 PS00008 MYRISTYL
N-myristoylation site

        152-157 GTGYSF
----
[5] PDOC00100 PS00108 PROTEIN_KINASE_ST
Serine/Threonine protein kinases active-site signature

        95-107 IIHRDVKPDNILL

Membrane spanning structure and domains:
```
 Helix Begin    End    Score Certainity
    1    185    205    0.727 Putative
```

FIGURE 2A

BLAST Alignment to Top Hit:

>CRA|103000001515936 /altid=gi|10946600 /def=ref|NP_067277.1|
          hypothetical serine/threonine protein kinase [Mus
          musculus] /org=Mus musculus /taxon=10090 /dataset=nraa
          /length=488
          Length = 488

 Score = 760 bits (1942), Expect = 0.0
 Identities = 371/399 (92%), Positives = 381/399 (94%), Gaps = 1/399 (0%)
 Frame = +1

Query: 103 VNFDHFQILRAIGKGSFGKVCIVQKRDTEKMYAMKYMNKQQCIERDEVRNVFRELEILQE 282
            VNFDHFQILRAIGKGSFGKVCIVQKRDTEKMYAMKYMNKQQCIERDEVRNVFRELEILQE
Sbjct: 89   VNFDHFQILRAIGKGSFGKVCIVQKRDTEKMYAMKYMNKQQCIERDEVRNVFRELEILQE 148

Query: 283 IEHVFLVNLWYSFQDEEDMFMVVDLLLGGDLRYHLQQNVQFSEDTVRLYICEMALALDYL 462
            IEHVFLVNLWYSFQDEEDMFMVVDLLLGGDLRYHLQQNVQFSEDTVRLYICEMALALDYL
Sbjct: 149 IEHVFLVNLWYSFQDEEDMFMVVDLLLGGDLRYHLQQNVQFSEDTVRLYICEMALALDYL 208

Query: 463 RGQHIIHRDVKPDNILLDERGHAHLTDFNIATIIKDGERATALAGTKPYMAPEIFHSFVN 642
            R QHIIHRDVKPDNILLDE+GHAHLTDFNIATIIKDGERATALAGTKPYMAPEIFHSFVN
Sbjct: 209 RSQHIIHRDVKPDNILLDEQGHAHLTDFNIATIIKDGERATALAGTKPYMAPEIFHSFVN 268

Query: 643 GGTGYSFEVDWWSVGVMAYELLRGWRPYDIHSSNAVESLVQLFSTVSVQYVPTWSKEMVA 822
            GGTGYSFEVDWWSVGVMAYELLRGWRPYDIHSSNAVESLVQLFSTVSVQYVPTWSKEMVA
Sbjct: 269 GGTGYSFEVDWWSVGVMAYELLRGWRPYDIHSSNAVESLVQLFSTVSVQYVPTWSKEMVA 328

Query: 823 LLRKLLTVNPEHRLSSLQDVQAAPALAGVLWDHLSEKRVEPGFVPNKGRLHCDPTFELEE 1002
            LLRKLLTVNPEHR SSLQD+Q AP+LA VLWD LSEK+VEPGFVPNKGRLHCDPTFELEE
Sbjct: 329 LLRKLLTVNPEHRFSSLQDMQTAPSLAHVLWDDLSEKKVEPGFVPNKGRLHCDPTFELEE 388

Query: 1003 MILESRPLHKKKKRLAKNKSRDNSRDSSQSENDYLQDCLDAIQQDFVIFNREKLKRSQDL 1182
            MILESRPLHKKKKRLAKNKSRD+SRDSSQSENDYLQDCLDAIQQDFVIFNREKLKRSQ+L
Sbjct: 389 MILESRPLHKKKKRLAKNKSRDSSRDSSQSENDYLQDCLDAIQQDFVIFNREKLKRSQEL 448

Query: 1183 PREPLPAPESRDAAEPVED-EAERSALPMCGPICPSAGS 1296
               EP P PE+ D +   D EAE +ALPMCG ICPS+GS
Sbjct: 449 MSEPPPGPETSDMTDSTADSEAEPTALPMCGSICPSSGS 487 (SEQ ID NO: 4)

>CRA|120000042903164 /altid=gi|13358640 /def=dbj|BAB33045.1|
          (AB056389) hypothetical protein [Macaca fascicularis]
          /org=Macaca fascicularis /taxon=9541 /dataset=nraa
          /length=368
          Length = 368

 Score = 733 bits (1872), Expect = 0.0
 Identities = 358/369 (97%), Positives = 361/369 (97%)
 Frame = +1

Query: 193 MYAMKYMNKQQCIERDEVRNVFRELEILQEIEHVFLVNLWYSFQDEEDMFMVVDLLLGGD 372
            MYAMKYMNKQQCIERDEVRNVFREL ILQEIEHVFLVNLWYSFQDEEDMFMVVDLLLGGD
Sbjct: 1   MYAMKYMNKQQCIERDEVRNVFRELGILQEIEHVFLVNLWYSFQDEEDMFMVVDLLLGGD 60

Query: 373 LRYHLQQNVQFSEDTVRLYICEMALALDYLRGQHIIHRDVKPDNILLDERGHAHLTDFNI 552
            LRYHLQQNVQFSEDTVRLYICEMALALDYL GQHIIHRDVKPDNILLDERGHAHLTDFNI
Sbjct: 61   LRYHLQQNVQFSEDTVRLYICEMALALDYLCGQHIIHRDVKPDNILLDERGHAHLTDFNI 120

Query: 553 ATIIKDGERATALAGTKPYMAPEIFHSFVNGGTGYSFEVDWWSVGVMAYELLRGWRPYDI 732
            ATIIKDGERATALAGTKPYMAPEIFHSFVNGGTGYSFEVDWWS+GVMAYELLRGWRPYDI
Sbjct: 121 ATIIKDGERATALAGTKPYMAPEIFHSFVNGGTGYSFEVDWWSLGVMAYELLRGWRPYDI 180

Query: 733 HSSNAVESLVQLFSTVSVQYVPTWSKEMVALLRKLLTVNPEHRLSSLQDVQAAPALAGVL 912
            HSSNAVESLVQLFSTVSVQYVPTWS+EMVALLRKLLTVNPEHR SSLQDVQAAPALAGVL
Sbjct: 181 HSSNAVESLVQLFSTVSVQYVPTWSREMVALLRKLLTVNPEHRFSSLQDVQAAPALAGVL 240

Query: 913 WDHLSEKRVEPGFVPNKGRLHCDPTFELEEMILESRPLHKKKKRLAKNKSRDNSRDSSQS 1092
            W HLSEKRVEP FVPNKGRLHCDPTFELEEMILESRPLHKKKKRLAKNKSRDNSRDSSQS
Sbjct: 241 WGHLSEKRVEPDFVPNKGRLHCDPTFELEEMILESRPLHKKKKRLAKNKSRDNSRDSSQS 300

FIGURE 2B

```
Query: 1093 ENDYLQDCLDAIQQDFVIFNREKLKRSQDLPREPLPAPESRDAAEPVEDEAERSALPMCG 1272
             ENDYLQDCLDAIQQDFVIFNREKLKRSQDLP EPLPAPE RDAAEPVEDE E+SALPMCG
Sbjct: 301  ENDYLQDCLDAIQQDFVIFNREKLKRSQDLPSEPLPAPEPRDAAEPVEDE-EQSALPMCG 359

Query: 1273 PICPSAGSG 1299
             PICPSAGSG
Sbjct: 360  PICPSAGSG 368 (SEQ ID NO: 5)
```

Hmmer search results (Pfam):
Scores for sequence family classification (score includes all domains):

Model	Description	Score	E-value	N
PF00069	Eukaryotic protein kinase domain	197.7	1.8e-55	1
CE00359	E00359 bone_morphogenetic_protein_receptor	9.8	0.044	1
CE00022	CE00022 MAGUK_subfamily_d	9.2	0.013	1
CE00031	CE00031 VEGFR	1.6	1.3	1
CE00292	CE00292 PTK_membrane_span	-92.7	0.0016	1
CE00291	CE00291 PTK_fgf_receptor	-116.1	0.041	1
CE00287	CE00287 PTK_Eph_orphan_receptor	-125.0	2.8	1
CE00286	E00286 PTK_EGF_receptor	-131.8	0.0056	1
CE00290	CE00290 PTK_Trk_family	-206.7	0.58	1
CE00016	CE00016 GSK_glycogen_synthase_kinase	-248.2	0.0061	1

Parsed for domains:

Model	Domain	seq-f	seq-t		hmm-f	hmm-t		score	E-value
CE00022	1/1	86	107	..	133	154	..	9.2	0.013
CE00359	1/1	95	144	..	272	326	..	9.8	0.044
CE00031	1/1	79	171	..	1051	1141	..	1.6	1.3
CE00286	1/1	3	212	..	1	263	[]	-131.8	0.0056
CE00287	1/1	7	212	..	1	260	[]	-125.0	2.8
CE00290	1/1	4	215	..	1	282	[]	-206.7	0.58
PF00069	1/1	2	228	..	26	271	..	197.7	1.8e-55
CE00292	1/1	5	233	..	1	288	[]	-92.7	0.0016
CE00291	1/1	1	233	[.	1	285	[]	-116.1	0.041
CE00016	1/1	1	303	[.	1	433	[]	-248.2	0.0061

FIGURE 2C

```
   1 GATGGAGCCC AGGCCACATG CTTGGGAAAG CGGGTGGTCC TGGGTGAGCC
  51 TGTCTGGCAG GAGGGTGAGG TCCTGTCTGG AGTAGACACC CGGTCTGTGT
 101 CCGCCGCACA TGTGACAGGG CCCAGAGGGC ACGGGAAGCC CAGGTGCCCT
 151 GTCCCCTGGT GAGTGGGCTG TGGGGGTTGA GCACCCCAAG GGAAGGCGTG
 201 TTCTTCCAGA GATTCCCCCG TAAGAGCTGA GCTGCATCGT GAGCAGGAGG
 251 GGCAGAGGGA GCCTGGACAG GGTTGGGTCG CCTGGGCAGG CAGCTTTCTG
 301 GGGTGGTGCA CAAGCGACAG AAAGCCTGAG GCTTGGGACC GGCTGAAGAC
 351 GGCCTTGGGC TGCGGGTGCT GGCGCCGCTC CAGGCAGAGG TGGGCTCTGC
 401 AGCCCCCCCC CCCCCGCCAG CTTCCCCCCC AGCAGCAAGA GCCCTGGCCC
 451 CAGCCAGTGG CCCCAAGCCA GGCTCCTGAG GGGCAGAGGG TGAGGGCCGG
 501 ATTTTCCACC ATATTTGTCT CACAGCCTGT CTGGTCCCCA CCCCAGGGCA
 551 ACAAACAGCC TTTCTGGAGC AGTTTCCAGA CCTGCAGTGG CCGCCTTGAG
 601 CCTGCAGTGA CCGTCTGCAG GAGGCCGCGG GTGCTGGGGC TGGCGCAGGA
 651 AAGCACCGTT GCTTCTGCGC CTGTGCAGAG TGAGGCTGGG GCTTCCATCC
 701 CGGGCACGGG ACTCCTCGGC CTCCTGCGGC CGTGTGCATG GGAGGAAGGC
 751 CGTGCTGCCG AGCCACGCGA CTCTGCCCCG TTGGCAGTGG GAAGCGGCAG
 801 GAGGGGGTCC TGCCAGGGGC AGCCAGGGGC TGCTGCAGCT TACGCTCACT
 851 GTCATTCTGA AACCCTCAAC TGGCTTTCAA AATAACAATT TAAAAAATGG
 901 TCATAGGAAA TGCAGGAAGT TCAGGAGAAA TCCCGCCCCG CCCCGCCCTC
 951 CCAAGGCGGC CTCGTTCAAG CTTAGTCTCC GTCTGTTCTC GGGCTGCCTG
1001 CAGCTGCCCC CGCTCCTCAG CAGGTGTGGC CGCGTGTTCA GGAGCCCCCA
1051 TCAGCACACG CGCCTCTGGG CAGCCCCCAA CACAAGGCTC TTCCTGTCCC
1101 TTCAGGCTCA GCTTTCCCCT CCCACCGGGC AGGGAGGTGC TGAGGCCACG
1151 CCTGTTGTCA GCTTCCTGGA GAGGCCCATAC TTAGAGCCCA TGGTGCCAGG
1201 CCAAGCCACT GTCTCCCGCA CCACAGTGCT GCAGCCGGAT CCACCCAGGC
1251 CACCGCTTGG CACCATCAGA CATGCTTTCT TAGTTTGGGC CAGCCCGGTG
1301 CTCTGTGCAC AGTGTGACTC CCAGTGGCCC CTTGCGGGAG GGAGGGTCAC
1351 CGCTTTGCTT CACAAAAGGA CTGAGTCTCA GGGAGGGGTC TCCAGTAAGG
1401 GCCCCGGAGC CAGGATGTGA CTGAGACAGG TGCCTCCAGG GCCACACACT
1451 AACCAGTACA GGAACCTCTG GGGGCAGAAT CATGGCCTCC AAAACACCCA
1501 CATCCAAATC CCAGAACAAA CATGTTACCA TTCGTGGCAG AGAGGAATGA
1551 AGATTGCAGA TTGAATTAAG GTTGTTAATC ATCTGACTTA ATTTTTTTAG
1601 AGACAGGGTC TCGCCCTGTC ACCCAGGCTG AGTTCAGTGG TACAGTCGTG
1651 GCTCACTGCA GCCTCAACCT CTCAGGCTCA AGCAATCCTC CCTTCTCAGT
1701 CTCCTGAGTA GCTGGGACTA CGGGTGTGTG CCACCACGCT TTCCGGGCAT
1751 GCAGCCAGGA GCCCAGGGCC ATCTGTGGCC CACCTTGAGA TCCAGAATCA
1801 TCCATTTCCT CCAGGCCCCC TGCTGGGCTC CAACTCCTTG AGGACCAGAG
1851 AGCAGAGGTT GTGGAAGGCC TTGGAAACGG GTCTGGATTA CCTGTCCTGG
1901 GAAGGTCTCT CCCAACCTGA GTGTCAGACA GGGGTTAGCT CTGCTGCTCA
1951 CAATTTTGTG CCTTAATTCC TGGCTTCCCT TTGGGGATCT TCATCCTCAA
2001 TTCTGATTGA CATCCTTGGC CACAAGGGAC CCCCCTGCTC ATTGATGCTT
2051 CTCACCCGTC ACCTCACTCT CATCCTCACT GCTAAGCAAT TAGCCGTGTG
2101 TTTGCGGCAT CAGTGTTGAC ACCGATGATC CATGCTCAGA GGGTACAGGC
2151 CTGAAGAGCT ATGTGGGGAC TGGCGCCCCG GAGGGGGTCC CGCTGTGGTG
2201 GCAGCGGTGG CCCCCAAGCC CCACGCTCAC TCTGTGTGTC TCCTTGCAGG
2251 ACAGGGTGAG GGGCCTTGGC CTCACGGTGT TGAGACGGGA GTCGGTCTCT
2301 GGAGTGTGGA GTGATGTGCG TCCAGGGTGA AGCTGCAGCC ATGGTTGATG
2351 AGGCTTCCTG AGCGCAGGGC TCTGTGCTCA GTGGGTTTTC GCATTCATTC
2401 CCCAGTACCC CTCCGGGCTG CTTGCTGCTC CAAGCCCTGG AAAAGGATGT
2451 TGGGGTTTAG GAAAGGCAAAA CTCCATGCCC AGGTCGTGGC TGGTGAGGGG
2501 CGCTCCCCGC ACAACCAGCC ACTGCTTGGC TCCAACTCAC GCCTGGATGC
2551 TGTTAGGCTG GACCCTGTCT GTTTGCAGAT AGCGCCTGTT GACAGATGTG
2601 TCCTGCTGCA GACTTGAAAC GCAGGACTGA GTCTCAGGGA GGGGTCTTCA
2651 CTAACAGCCA TGGAGCCAGG ATGTCACTGA GACAGGTGCC GCCAGGGCCA
2701 CACACTAACC AGTACAGGAA CCTCTGGGGG GCAGAATCAC GGCCTCCAAA
2751 ACACCCACAT CTTAATCCCC AGAGCAAATA TGTTACCATA TGTGGCAGAG
2801 AGGAATGCAC CTGCCGCGTC TAGGAGGCAG GGGGCTGCG GCGCGTCCCC
2851 AGGTGTCCCC TTGTGTCCTG ACGATGCGTC CCCAGGCATC CCCATGCATT
2901 TCCAGGTGTC CCCACGTGTC CCCAGGCGTC CTCAGGCGTC CTGCCGAACA
2951 GCCCTGTGCC CTCCAGGTGT GCATTGTGCA GAAGCGGGAC ACGGAGAAGA
3001 TGTACGCCAT GAAGTACATG AACAAGCAGC AGTGCATGGA GCGCGACGAG
3051 GTCCGCAACG TCTTCCGGGA GCTGGAGATC CTGCAGGAGA TCGAGCACGT
3101 CTTCCTGGTG AACCTCTGGT GAGCCTGCCA TGGCCTCGCT GCAGAAAGAC
3151 TTACCGTCCT GAAGCCGGGA AGGCACTGGC TATCCTCTCC TGCCCTTGGT
3201 GCTCCTTGGC CCTCGTGCCT CGGAGATGCC TCTGCCACCG ACAGGCCCTC
3251 TCTAGCCCTC CTCATCTATC CCCCTTCCTG TGCCCCAGGC CTGGCAGTGG
3301 CCCAGGTGGC CATGACATGC TGGGGTTGGT TAATGCAGTG TCTCTTCTGA
3351 GCCTGCCGGA AGACCAGGGC TTCCCTACAA TGGAGATGTG CTCCCATGGA
3401 GTCTCTGGCA CTAGTCAGAG AGGGAGAGAG TTTAGGGACT GAAAAACTCA
3451 CCACTGTCAT CACCATCACC ATCACCACCA TCATCGCCAT CACCACCACC
3501 ATCACCACTG CAACCATCAT CACTCTCATC ACCATCATCA CTATAATCAA
3551 CACAATCACT ATTGTCACCA CCATTACCAC CACCACCACG ACCACAATCA
3601 CTGTTATCAC CATCACCACC ACCCTCAGTC CTCACTACCG TCATCCTCAC
3651 CATCACCGTC ACCACCACCA CCATCACTGC CATCGTCAAC ACCATGGATG
```

FIGURE 3A

```
3701 CTGGTTGTTA AATGCCAGCT CTTTGCCCAA CACTGTCAAG AGTGGTACCT
3751 ACACGGCCTC ATTTTCTGTA ACAACCCTCC GAGGCAAATG TCTGTATCCC
3801 CATTTTACCG AAGAGGAGGC CGGGCAGCCT GAAGCACCCG GAGCTGGCAC
3851 TGTAGCTCTG CTCTGCATTT GCCACTCCCA GGTGCCTCTG GCCCCAGCTG
3901 GGCCACCTCC AGCACAGGGT GGTGTGTCTT TCCTCAGGAT CTGGGCTCAG
3951 AGCTGCTCTG GGCTGGGGTG CAATCAGTGC CTTGGGCAGG CCCCTCCTCC
4001 TGGGAATGCC TGGTGGCTGA TGCTGGGGTG GGGCTGTGGT CCTTAGGGGG
4051 AGTGTGTCAG CTGTGGGAGC AGCCATGACT GGCTCCCCAG CTGTGCGCAC
4101 AACAGGCCTT CCATCGGTGC CCACAGGTAC TCCTTCCAGG ACGAGGAGGA
4151 CATGTTCATG GTCGTGGACC TGCTACTGGG CGGGGACCTG CGCTACCACC
4201 TGCAGCAGAA CGTGCAGTTC TCCGAGGACA CGGTGAGGCT GTACATCTGC
4251 GAGATGGCAC TGGCTCTGGA CTACCTGCGC GGCCAGCACA TCATCCACAG
4301 GTGTGTGCGT GGCAGACGGC GCAGGTACCT GCTGAGGTGG GCGGGGCTGA
4351 AGCAGCCTTA GGTCAGGCTG CCGGCACGGC GGCCGTACTC CCTCAGAGCG
4401 GGTCTAGCTC CTCTGCCCCA CCCTTGCCTG AGTGCCTGCC CCCAGCTGTG
4451 GCACCTGTGC CGACCAGGTC AGCCCCATAG CTGTGTGCCT GGTGTCCATC
4501 TGGGGGGACC TCGTCCCAGC AGCCCCAGCT GAGACTGGGC ACAGTGGGCT
4551 GTTAGCCCTG GTGGACAGAC CACCAGGCTG GGTCACAGCA GGTGGCCTTC
4601 ACCTGGTCCA TTTAACTGAA GACTCCTGTT TGCCCATCCA ACACATCCCA
4651 GGGAATCCAA ACTAATTTTA ACATTAGCTT AAAGCAGATG AAATTAGGAA
4701 GCAGAGCTGG TGTGATGGCT CTGAAAATAA AATTTAAAAA AAGAAAATAG
4751 GAAGCAGATT ATGAAGGAAG TGAAATTGGG AAGCAGAAAT TAGGCTGAAA
4801 TTCCGCAGCA ATGGAACAAA ATGAAAATAT CTGTGAGGTA TATTTTAAAG
4851 TCGAATGGAC TGGTGTTTGC ATTTCTGCTC TTGGGGACTC GGATGTCTGA
4901 TTATGACCTA GGCACCAGTC ACTGAGCACT GGCTGTGTAC CTGGAAAAGT
4951 TGGGACAAAG CAAGAGCCGA GGTGGCTTGG TCTCCTAGAG GCCGAGTCTT
5001 GGAGGGGGAG GGCAGACCCT GCCAGCAATT GCTCTCGTCC TCTGGGGCTC
5051 CAGGCCCCCT CCCAGCATCT GGTGCCAGGT GTGTGCTGCT GCCCAGATGC
5101 CACAGGGAAC GAAGTGGCTG ACTTCATCGC CTCTGCCCCC ACGCAGGGGT
5151 GTGAGGTCCT AGCATCATCC AAGGACCAAG TCAAGCTCCC AGGCCTCTGC
5201 CTCGAGTGGG TTGGTGGGAT GTCCTGGGGA CTCCAGGGAT TGTGACAGAG
5251 ATTCCAGGGC AGAAACAGGG CAGATTCCCA ACTCACGTTC CCACTTTCTG
5301 CTCTTTCTAG AGATGTCAAG CCTGACAACA TTCTCCTGGA TGAGAGAGGT
5351 GTGTGGGGTT GGGTGTGGGC AGCCCAGGTG GGTGGTGGCA GGGATGGGCC
5401 TGTCAGGGGA GGAGGATCCT GCACGCAAGG ATGCATCTCT GGTCCTGGGA
5451 CAGCCACACC TGACCCCTCT CTGCACAGGA CATGCACACC TGACCGACTT
5501 CAACATTGCC ACCATCATCA AGGACGGGGA GCGGGCGACG GCATTAGCAG
5551 GCACCAAGCC GTACATGGGT GAGCCCGAGC TGGGGTTCCA GATGGGAGCT
5601 GGCTTCCTCC AGGTGGGAAG GACAAGACCT CGGTGGCTTC TCTGTCCCAC
5651 CCTGGAGGCA GCCTGGTCTC GGGATGTGGC CTCAAGGTGC CGGCCCTGTG
5701 CCCACGGGTC CGGGCTGTGA CCCCGTGGCA GCTGTTTTTC CTTCTTTCTG
5751 TCGGAAAGCT CCGGAGATCT TCCACTCTTT TGTCAACGGC GGGACCGGCT
5801 ACTCCTTCGA GGTGGACTGG TGGTCGGTGG GGGTGATGGC CTATGAGCTG
5851 CTGCGAGGAT GGGTATGGAC CCCCTGCAGC CCCCGGGCTT GGCTGCCAGG
5901 CCCCTGCTCT CTGCCCCCAC CAGTGCTGGG GAGGGGGTGG CTGCCCCAGT
5951 GCCCAGGTGC GCAGGGATGC CTCCACTGTG TCTGAGGAGT CACGCTTTTA
6001 TCGAAGTGTG TAGTTGGTGA TGGAATGCCT GAGCAGGAGG AGGAAGGACA
6051 GACTCACTGT GGTTTCCCGG GGCCGCTGCT GGTGCCTGCA GGCCAGCCTC
6101 TGTGGGGGTG GACAAGGCTG AGAACTGGCC AGCAGGGGTG CTGCCTCGGA
6151 ACTTTCCACA AAAAGTTTCT TTTGGGGCCC TGTGCTCTTA CCCTTGTGGC
6201 CACGGCGAGG CCAGTCCTGG AGACCGGGAG GCTGGGGGTC CTCTTGTGGA
6251 CCGTACCCCT CAGCCCTGCA CAGGACCCCA CCTCTGAGGA AGCCAGCTCC
6301 CTCCTGGCCC TCTGGGGCTG ATCTACCTGG ACCCAGGCCG CCTGGGATCC
6351 CAGCCAGATG GGCGCAGCAG CCAGGGCGCA GGACCCAGGC GTAAGCTTTA
6401 TCTCACCCAG GCTCCTCCGC GGCAGGTGGA GGCCAGGCTG TGCTCAGAGC
6451 TGTGCCTGCA CTTGGGGTGG GGGGAGGGGG TCCTCTCAGG GCGATGGCAC
6501 CTGTGTGTCTGG CATTGTTCTG GGTGTCCTGG GGGCCAGGAG GACCTGCCCA
6551 GCACTGCCTC CCTGTCTCCA GAGGCCCTAT GACATGCACT CCAGCAACGC
6601 CGTGGAGTCC CTGGTGCAGC TGTTCAGCAC CGTGAGCGTC CAGTATGTCC
6651 CCACGTGGTC CAAGGAGATG GTGGCCTTGC TGCGGAAGGT GAGCCCCCAT
6701 CCCTGAGCCT CCTCACCCTC CGAGCACCCA CCTCCCTACC TCACTTACCT
6751 GCGGCTCGGG ACACCCCCTC CAGTGCACAG TTAGTGCCGC TTCCTGGCAG
6801 GCACAGATCC CTTCACTGCA ACCTGTGGGG GCCTCCGCAG ATGGCAGCCC
6851 CAAGCCCCAG GAAGCGAGCT GGTGGCAGGC TCTGTGGCCC TCTCATGGCG
6901 AGCCCTACCA GGTCACTGGT GTCCTGGTGA TCCCCTGAGC TGCGTCTCCA
6951 GGCACCCTCA CAGCAGGCTT GTGCGCCCGC CCTGCCGGTC ACCACAGAGG
7001 AGCCCTGAGC CACGTCCCCT TGTCCCACGT ATTGGAAGGA GGGTGTCGGG
7051 TGGGTGGGTT CAAGCCCATG CTATCTCCGG GACCCTTTGC CCCAGTGCCT
7101 CCTGGGGAAG GTGGGCAGCC ACTGCCCACT GCAAACACCT CTCAGGGGAC
7151 ATCAGCCTGG CAGGGACACG GGCAGGGGTG TGGCCATCAG TAGTGCCTCC
7201 CATTTGTGAT CTGGTGCTAG GCTGGGGCTG TGCCTTGACT GGGCTGCACA
7251 GCTCCCTGTT CCGAGCCTCG CATTAACACC ATAGGGGGTT CGGAGTCAGA
7301 GCCAGGCCCA GGCAGGACAG GGAGGGGAGT GAGTGTGCCA CACGGGCCCG
7351 GCTGCCTCCC GGCCCCCGTG TCTCAGGCAG GTGGGGCCTC CTGCCCTGGA
```

```
7401  ATTGTAGCCA  AGCAGCCTAA  AGCCTTGGGG  AGGCCTTGCC  TGCCGGGGCC
7451  TCTCCCCAGC  CCCGAGAGTC  TCTTAACTCT  GCTGTAGCCC  CATGAAGCTC
7501  AGTCACACCT  GCCCAGGTGG  CTCACAAGGT  GGCACTGGGC  TAGAGAGGGC
7551  CTGCGTGGGG  ACTGGGGATG  ACCCACACGC  CCAAGCCCAG  GTCTGGGAAA
7601  CCTCGCACGG  GGTCTGGGTC  TGCGGCATTT  TCCCTGGAAA  GGCGGGAGGT
7651  GCCAGCGCTG  GGATGTTGCT  TCCCAGGCCA  TGCATGGCTG  CCCCGGGCTC
7701  ATCTGGCCTG  TGGAGGTCCC  ATGATTCGGT  GAAGGAAGTG  GCTCTGGGAT
7751  AGTTACTGTG  AGGCCAGCCA  TGTGCCGAGT  GTTAGCCGCT  AGCCGGGCCT
7801  CGGCTGCCAC  CTCCTGGCAA  ATCCCAGCAG  AGCCTTCCCT  GCAGATCCCT
7851  CTGCTGTCCT  CTGGCGCCAG  GGGTTTAGGT  AGCAGCACTG  AGAACAGGCG
7901  TCCCTTGGGC  CACATGCTGA  GCCAGCCACG  GTGCTTGTCC  TGATGTCGGC
7951  CGTCGGCACC  ACCCTTCCTC  GCGTGGCCCT  GAGGTTCCTG  AATTCTGAAC
8001  CTGAGGCTTG  GTGGGACCCT  CCTCAAGGTG  CCCTGGCCTG  GGGGTGGCGG
8051  GCTATTCCGT  GCTGGTGGGC  TGTGGGCCCT  GGACCCTCTG  ACTCATGCCT
8101  GGTTGCAGCT  CCTCACTGTG  AACCCCGAGC  ACCGGCTCTC  CAGCCTCCAG
8151  GACGTGCAGG  CAGCCCCGGC  GCTGGCCGGC  GTGCTGTGGG  ACCACCTGAG
8201  CGAGAAGAGG  GTGGAGCCGG  GCTTCGTGCC  CAACGTAAGC  CTGTGGGCGG
8251  CTCAGGTGGG  GGGCCCTGGG  GATGGATGTG  GCGTCCTCCA  CGGGCCGGGG
8301  CTCAGCACCC  ATCCCTCTGT  AGAAAGGCCG  TCTGCACTGC  GACCCCACCT
8351  TTGAGCTGGA  GGAGATGATC  CTGGAGTCCA  GGCCCCTGCA  CAAGAAGAAG
8401  AAGCGCCTGG  CCAAGAACAA  GTCCCGGGAC  AACAGCAGGG  ACAGCTCCCA
8451  GTCCGTGAGT  GCCAGGGCAG  GCTCAGGGCG  CGGCGGCGGG  CTGGGCTTGG
8501  GGCTCCTCTC  TACCACCGAG  CAAGGTGTGT  GGGGACCCCT  GACAGTGCAC
8551  ACGTCTCGGA  AGTCCAGCAG  ACCGTTTCCT  GAAGTCCTGA  GAAGGCCAGA
8601  GACCTCCCTT  CTGCCTTTCC  CAGCCCCCAC  CTCGCTCCTT  ATGAAGCAGG
8651  TGGGCAGGGA  CAACCAGGGC  TGGGGTTATG  AGTGCACGGG  GATGGCCATG
8701  TGAAGCCTTC  GTGCTTGCCC  AGGTGTGCTG  GTGTTGGTTG  TGTGTGCGGG
8751  GACGGCTATG  TGAAGCCCTC  ACACTCGCCC  AGGTGCGTCG  GCATCAGGTA
8801  TGTGTGCCGG  GACAGCCATG  TGAAGCCCTC  ACACTCACCC  AGGTGCGTCG
8851  GCATCAGTTG  TGTGTGTGGG  GACGGCCATG  TGAAGCCCTC  ACACTCGCCC
8901  AGGTGTGCTG  GCTTTGGTTG  TGTGTGCAGG  GATGGCCACA  TGAAGCCCTC
8951  ACACTCGCCC  AGGTGCGTCA  GCATCAGGTG  TGTGTGCGGG  GACGGCCATG
9001  TGAAGCCCTC  ACACTCGCCC  AGGTGCGTTG  ATGTTGTGTG  TGCAGGGATG
9051  GCCATGTGAA  GCCCTCACAC  TCACCCAGGT  GCGTTGATGT  CAGTTGTGTG
9101  TGCAGGGACA  GCCATGTGAA  GCCCTCAGAC  TAGCCCAGGT  GTGTCGGTGT
9151  CAGTTGTGTG  TGTGGGGATG  GCCACGTGAA  GCCCTCACAC  TTGCCCAGGT
9201  GCGTTGATAT  TAGTTGTGTG  TGCAGGGATG  GCCACGTGAA  GCCCTCACAC
9251  TCACCCAGGT  GCGTTGATGT  CAGTTGTGAG  TGTGCGCAGG  GATGGCCACA
9301  TGAAGCCCTC  AGACTCGCCC  AGGTGTGCTG  GCTTTGGTTG  TGTGTGCAGG
9351  GACGGCCATG  TGAAGCCCTC  ACACTCGCCC  AGGTGCGTCA  GCATCAGTTG
9401  TGTGTGTGGG  GACGGCCATG  TGAAGCCCTC  ACACTCACCC  AGGTGTGTCG
9451  ACATCAGTTG  TGTGTGGGGG  GACGGCCATC  TGAAGCCCTC  ACACTCACCC
9501  AGGTGTGTCG  GTGTCAGTTG  TGTGTGCGGG  GATGGCCACG  TGAAGCCCTC
9551  ACACTTGCCC  AGGTGCGTTG  ATATTAGTTG  TGTGTGCAGG  GATGGCCACG
9601  TGAAGCCCTC  ACACTCACCC  AGGTGCGTTG  ATGTCAGTTG  TGAGTGTGTG
9651  CAGGGATGGC  CACGTGAAGC  CCTCAGACTA  GCCCAGGTGT  GCTGGCTTTG
9701  GTTGTGTGTG  CAGGGACGGC  CATGTGAAGC  CCTCACACTC  GCCCAGGTGC
9751  GTCAGCATCA  GTTGTGTGTG  TGGGGACGGC  CATGTGAAGC  CCTCACACTC
9801  GCCCAGGTGC  GTCAGCATCA  GTTGTGTGTG  TGGGGATGGC  CACGTGAAGC
9851  CCTCAGACTA  GCCCAGGTGC  GTCGGCATCA  GGTGTGTGTG  CCGGGACAGC
9901  CACGTGAAGC  CCTCACACTC  GCCCAGGTGT  GCCGGCTTTG  GTTGTGTGTG
9951  CGGGGACGGC  CACGTGAAGC  CCTCATGCTC  ACTCAGGCAT  GCTGGTATTC
10001 TGGGGCTGCC  AGGACAGGTG  ACCACGAATC  AGGTGGTTGA  AGAACAGCAA
10051 TGCGTCTCTC  TGAGAGGATC  TGAGTCGTAA  TGAAATGGTC  TCCTTCACAG
10101 CCGGCTGTGC  GTGAACTACT  CTGTCTCCTG  CAGCTCCCCT  GTCTTGATAA
10151 TTGGCTGTCT  AGGCAGCGGG  TAAGGTGAAC  CCCTTGGGCA  GTTATGTGAT
10201 GATCTCAGTT  TCTGTAAACC  GGAAGTCCAG  GCATGGTGCA  GCTCTGTTCC
10251 CTGCTTCGGG  GTCTCACCAG  AATGTGAGCT  AACATTGAGG  TCGTGGCCTT
10301 GTCAGGTGCA  GCTCTGTTTC  CTGCTTCTGG  GTCTCCCCAG  AGTGTGAGCT
10351 AATATTGTCT  GAGGTCGTGG  TCTCATCAGG  GATTTGACAG  GTGCTGTGGT
10401 TGAAATGTTT  CCCTTAAAAC  TCGTGTTGGA  ATTTGCTTCC  TATTGTGATG
10451 GTGGTAGAGA  TGGGACTTTT  GGGGGCTGAT  GGGGCCACGT  AGGTTCTTCC
10501 AGCATGGATG  GGGTTAATGC  TGTTGTAGAA  GGGTGACTTT  AGTCCTCTTT
10551 TGAGTCTTTG  ATCCTCTGCT  ATGTGAGGAC  GTGGTGTTCC  CAATGTGGAC
10601 GTGGTTCGTG  TTCCATGTGA  ATGTGATATT  CACAATAGAG  CATCAACAGG
10651 CTCCCTTTTA  ATCAGCAGAT  TTAAAAAGAA  ATGTGTTGTC  TCATGGCTTG
10701 GAGGCCTGAG  TCCAAAGTTA  AGATGTCAGC  AAAGCCGTGC  CCCCTCTGAA
10751 GGCTCTCCGG  GGAGGAAAAC  CAGTCCTTGC  CCCTCTACCC  TCCGGTAGAG
10801 GCTGCCTTGG  CCTAGACGCA  TCCCCCCAGC  CCCTGCTTCG  CTGCCGCGTG
10851 GGGTCGGCCT  GTGTGTGCGT  CTCCATCTCC  TCCCCTCTTC  TCATAAGGAC
10901 ACCAGGCATT  GGATTTAGGG  CCCACCCTGA  TCCAGTATGG  CCCCATCTTA
10951 TCTTGATGAT  ATCTGCAAAG  ACCTCACTTC  CAAATGAGGT  CACATTCACA
11001 GGTACCCAGG  ATTAGAATTT  GAGTGTGTCA  TTTTTGGGGA  CACAGTTTGG
11051 CCCATACCAC  CAGGATGTGG  CTGATATTCA  CCAAGGAGTA  GCTATGGTTG
```

FIGURE 3C

```
11101 TGTGTTGATG TCAGGGTGAC GGTGATGACC CTGGGTCCCT CGGTGGTCCC
11151 CTTGCCCCTG AGTCTGCCTG AGCCTGTGGT GGATGTCCTG GGAAACTCTT
11201 GTGCCTCAGC CCCCGTGCAG CCTCCTCAGA CCTGGTGGGC CCTGTGTTGC
11251 TCCTGGGCAG AAGACGGGTG TCAGTCCCCT CCTCACCATG ATGTGGGGGG
11301 CAGGGGTGGG GTCATGCCCT GGGTGCCCTG ATTTTGGGGG GAACACGGCC
11351 CCCCAGTGGG TCAGGCTCCC ATCCTCGCCC CTCCTCCAGG ACGGCTGCCG
11401 GCAGCCCTGG GTGTCCTCAG GCAGAATTGC TGGTGGAGAG CTGCTGTCTG
11451 CCAGGTGGCC ACTGTGAGGC ACTGCTGAGA GCCACAGGAT GGTTGGAAGG
11501 TTCTCGGGGT TGGGGGTTCT TTGGCATTGC CCCCATTGGA TGTTTAAGTT
11551 TTCCCTACCA GAGCATGTCC AGAGCCAGGG CTCTGGGGTG TAGAAACAGG
11601 CCCAGGATGA GTTAGGAACC CTCATGGGAG ACTCAGGGAT GGACAGTGTG
11651 CAGAGCCCAG CTGGCCATGC TGAGTTCCCA GGAGGCTCTG GCTGGGAACA
11701 GGTAAGGCCA GGCACCTGTG AGCGGGAGGA GCTCGGCTTT GTCTTGGGTT
11751 GGTTGTGTGG AGATGTTTTG GCTTGAGGGT AGGAGGTGTT CTGAAAGGAA
11801 AGCATCACTC CAAAAAAAAA GTCCCACTGT TAACCTTGAG GCTGAGAGAG
11851 GTTTTTTGGAA ACAGCTTTAT TTTGATATAA TTCACATTCC ATGCAATATA
11901 CAGTGCATCC ATGTAAAGCA TATAATTCCA TGGTTTTTAA TATAGTCACA
11951 GGGCTGTGCA TTCTCCACCA CAATCTGATT TTAGAACCTT TTCATGTAAT
12001 GTAAGAGAAA GACCCCACCT ATTAGCAGTC ATGCCCCATT CCCCTCTTCT
12051 CCCCTCCCCT GGCAGCCACG AAGCTACTTT CCGTCTCTGT AGGGTTGCCT
12101 GTTGTGGGCG TTTCATGGAA GTGGAGTTAC ACACTATGTG GTCTTTGCAG
12151 CTGGCTTCTT TCACTTCGCA GGATGCTTTT GAGGCCCGTC CACGTTGTAG
12201 CCTGTCAGTG CTTCATTCCT GTTGATGGCT GAGTAATATT CCACATATGT
12251 ATCACCTTTC CTTTATCCAG TCATCAGTTG ATGAGTATTT GTTCTTTCCA
12301 CTTGTTAACA TTTTTCATTA TCATGAATAA TGCTGCTGTG AACATTCACA
12351 TACAAGTCTT TGTGTGGATA TGTACTTTTA TTTTTGGGGG GCACATACCT
12401 AGGCATGAAC CCGCTGGGTC ATATGTGACT CTGTGCTTCA TGTTTGGAGG
12451 AAACACCTAC CCTTTTCTAC AGCAGGTGTG CCATTTTATG CCCCTACCAG
12501 CAGTGTGTGA GGGTTCTAAT TTCTCCATAT ATTTACCAAG TCCTGTTATT
12551 GTCTGGTTTA TTTTTTTAAA AATCATAGTT ATCTTAGTGT GCAGTGATTG
12601 TGTGGTTATG GTTTGCATTT CTCTGATGAT ATTGAACATC TTTTGAGGTG
12651 TTTTATCAGG CATTGTGTCT AGAGAAATGT CTATCCAAAT GTTTTAAAAT
12701 TTTTATTGTC TTTTTTTTAG TCTACTCTGA CAATATATTT TAATTGGCAT
12751 ATTTATTTTA CTTTATTTTT TTTTAGAGAC AGGGTCTTGC TTTATTACCA
12801 AAACTGGAGT GCAGTGGTGC AATGAAGGCT CACTTCAGCC TTGACTTCCT
12851 GGGCTCAAGT GATCCTCCCT GCCCCAGCTG CCAGAATGGC TGGAACTGTG
12901 GGTGTGCACC ACCACACCTG GCTCATTTGA AAAAAATTTT GTTGTAGAGA
12951 CAGGGTCTCA CTATGTTGTC CAGGTTGGTC TCAAACTCTT GGAGTCCTCC
13001 CACCTCAGCC TCCCAAAATG CTGGGATTAC AGATGTGAGC CACTGTGCCT
13051 GACCTAATTG GTGTATTTTA GACCATTCAC ATTTAAAGCG ACCAGGGAGG
13101 CTGAGGCAAG AGGACTGCTT GAGTCCAGGA GTTTGAGACC AGCCTGGGCA
13151 ACAAGGTGAG ACCCATCTGT ATTAGTCTAT TTTCACACTG CTAATAAAGG
13201 CATACCTGAG TCTGGGTAAT TTATACAGGA AAAAGGTTTA ATGGATTTAC
13251 AGTTCCACAT GGCTGGGGAG GCCTCACAGT CGTGGAAGGC AAGGAGGAGA
13301 AAGTCACATC TTACATGGAT GGCGGCAGGC AAAGAGAGAG CTTGTTCAGG
13351 GAAACTTTTG TTTTTAAAAC CATCGGATCT CATGAGACTC ATTCACTATC
13401 ATGACAACAG CACAGGAAAG ACCCGCCCCC ATAATTCAAT CACCTCCCAC
13451 CAGGTTCCTC CCACAACATG TGGGAATTGT GGGAGTCACA ATTCAAGCTG
13501 AGATTGGGAT GGGGACAGAG CCAAACCATA TCATTCTGCC CCAGCCCCTC
13551 CAAAATCTCA TGTCCTCACA TTTCAGAACC AATCATGCCT TCCCAACAGT
13601 CCCCCCATAGT CTTATTTTGG CATTAACTCA AAAGTCCACA GTCCAATGTC
13651 TCATCTGAGA CAAGGTAAGT CCCTTCTGCC TATGAGTCTG TAAAATCAAA
13701 AGCAAGTGAC TTCCTAGATA CAATAGGGGT ACAGGCATTG GGTAAATTCA
13751 GCCATTCCAA ATGGGAGAAA TTGGCCAAAA CAAAGGGGCT ACAGGCCCCA
13801 TGCAAGTCTG AAATCCAGCA GGCCTGTCAA ATCTTAAAGC TTCAAAATGA
13851 ACATCTTTGA CTCTATCTCT CACATCCAGG TCATGCTGAT GCAAGAGTTG
13901 GGTTCCCATG GTCTTAGGCA GCTCTGCCCT TGTGGCTCTG CAGAGTACAG
13951 CCTTCCTCCC GGCTGCTTTC GTGGGCTGGC ATTGAGTGTC TGTGGCTTTT
14001 CCAGGTGCAT GGTGCAAGCT GTTGGTGGAT ATACCATTCT GGGGTGTAGA
14051 GGATGGTGGC CCTCTTCTCA GAGCTCCACT AGGCAGTGCC CCAGTGGGGA
14101 CTTTGTATAG GGGCACCAAC CCCACATTTC CCTTCTGCAT TGCCCTAGCA
14151 GAGGTTCTCC ATGAGGGCCC CACCCCTGCA ACAAACTTCT GCCTGGACAT
14201 CCAGGTGTTT CCATACATCC TCTGAAATGC AGGCAGAGGC TCCCAAACCT
14251 CAATTCTTGA CTTCTGTGCA CCTGCGGGCT CAACACCACA TGGAAGCTGC
14301 CACAGCTTGG GGCTTGTACC CTCTGAAGCC AAGGCAGCGA CTGTACCTTG
14351 GCCCCTTTCA GTCATGGCTG GAGCAGCTGG GATGCAGTGC AGCAAGTCCC
14401 TAGACTGCAC ACAGCAGAGG GACCCTGGAC CTCGCCCATG AAACCATTTT
14451 TTCCTCCTAG GCCTCTGAGT CTGTGATGGT AGGGGCTGCC GCAAAGGTCT
14501 GTGGCATGCC CTGGAGACAT TTTCCCCATT GTCTTGGTGA TTAATATTCA
14551 GTTCCTTGTT GCTTATGCAA ATTTCTCCTC AGAAAATGGG GTTTTCTTTT
14601 TTTTCTCTCT TTTTTTTTTT TTTTTGAGAC AGTCTTGCTC TGTCACCCAG
14651 GCTGGAGTGC AGTGGTGCAA TGGCGGCTCA TTGCCACTGC AACCTCCGCT
14701 TCCTAAGTTC AAGTGATTCT CCTGTCTCAG CCTCCCAAGT AGCTGGGATT
14751 ACAGGCACGC ACCACCACAC CCAGCTAATT TTTGTATTTT TAGTAGAGAA
```

FIGURE 3D

```
14801 GGGTTTCACC ATGTTGGCCA GGCTGGTCTT GAACTCCTGA CCTCAGGTGA
14851 TCTGCCTGCC TTAACCTCCC AAAGTGCTAG GATTACAGGC GTGAGCCACC
14901 GTGCCCAGCC AGGAGTTTCT TTTCTATTGC ATTGTCAGGT TGCAAATTTT
14951 TTGAACTTTT ATGCTGTTTC CTTTTTAAAA TGGAATGCCT TTAACAGCAC
15001 CCAAGTCACC TCTTGAATGC TTTGCTGCTT AGAAATTTCT TCTGCCACAT
15051 ACCCTAAATC ATCTCTCAAA TTCAGAGTTC CACAAATCTC TAGGGCAGGG
15101 GCAAAATGCT GCCAGTCTCT TTGCTTAAAG CATAACAAGA GCCACCTTTG
15151 CTGTAGTTCC CAACAAGTTC CTCATCTCCA TCTGAGACCA ACTCAGCCTG
15201 GACTTCATTG TCCATATCAT TATCAGCATT TTGGTCAAAG CCATTCAACA
15251 AGTCTCTAGG AAGTTCCAAA CTTTCCCACA TTTTCCTGTC TTCTTCTGAG
15301 CCCTCCAGAT GGTTCCAGCC TCTGCCTATT ACCCAGTTCT AAAAAGTTGC
15351 TTCCACATTT TCAGGTATCA TTTCAGCAGC GCCCTACTTT ACTGGTACCA
15401 ATTTACTGTA TTAGTCTGTT CTCACGCTGC TAATAAAGAC ATATCCGAGA
15451 CTGGGAAATT TATACAGGAA AAAGGTTTAA TGGAGTTACA GTTCCACATG
15501 GCTGGGGAGG CCTCACAATC ATGGCGGAAG GCAAGGAGGA GCAAGTCACA
15551 TCTTACATGG ATGGCAGAGA GAGCTTGTGC AGGGAAACTT TTGTTTTTAA
15601 AACCATCAGA TCTCATGAGA CTCATTCACT ATCATGACAA CAGCACAGGA
15651 AAGACCCGCC CCCATAATTC AATCACCTCC CACTGGGTTC CTCCCATGAC
15701 ACACGGGAAT TGTGGGAGTC ACAATTCAAG CTGAGATTGG GGTGGGGAGA
15751 CAGCCAAACC TTATCACCAG CTCTATAAAA GACAAAAAAA TTAGGCAGGC
15801 ATAACAGTGC ATGCCTGTAG TTCCAGTGAT GTGAGAGGAT TGCTTGAGTC
15851 CAGGAGTTTG AGACCAGGCT GGGCAACATG GCGAGACCCT GTCTCTACAA
15901 AAAAAAATTA TCTGGGTGTG GTGGGATACA CCTGTGATAC CAGCTACGCA
15951 GGAGGCTGAG GCAGTAGGAT TGCTTGAGCC CAGGAGTTCA AGGCTGCAGT
16001 AAGCTATGAT CATGCCCCTG CACTCCAGTC TGGGTAACAG AGAGACACGC
16051 TGTCTTGTAA ATAAATAAGT GGTCATTTAT ATAGTTCAAT ATGATATCTA
16101 CCTTATTTGT AACTGTAGTC TATTTATTGG TCTTCCTTTT TCCCTATTTT
16151 TCTGCCTTTT CTGGTTTTAA TTAAGCATTT TATATTATTC TAGTTTATCT
16201 CCTCTCCTGG CCTGTTAATT ATACTTCTTT TTGAAATATT TTTAGTGGTT
16251 GGCCTGGACA TTGCAGTATA CCTTTACCAT ACAGTCTACC TTCACCTGAC
16301 ACTCTGCCCC CTCATGTGCA GTGGGATGCC TTGTGACGGC ACCTCTCGTG
16351 CAATCCTCCT GTTCCTGATG ACATTGCTGT CATTCATTTC ATTTATCTGT
16401 ATGCTATAAT TGCTCATTAC ATTGTTACTA CTGTTATTTT AAACAGTTAT
16451 CTTTTGGATC AATTAAGAAA AATTAAAAAT TTCATTTTAC CTCTATTCAT
16501 TCCTCCTCTA AAGTGCTTCC TTTCTTTATG CAGACCCAAG TTGCTGACCT
16551 AAATCATTTT CCTTTCCCCT GAGGAACTTC GTTTAACATG TCTTATAGGA
16601 CAGGTCCAAC AGAGATGAAT TCCTTCCCTT TTTGTTTGTC CAAAAAAGTC
16651 TTTACTTTCA CCTTTAAAGA ATAATTTCAC TGGATATAGA ATTCTAGATT
16701 GGTAGGTTTT TTACTTTCAA CACTTTAAAT ATTTCACTCC GCTCTCTTCT
16751 TGCTTATGTG ATTTCTAACA GGCAGTCTGC TCTAATTCTT TTTCTGTAAG
16801 TAGAATTCCT CGCCACCCCC CACCCCCAGC TCATTTCAAG ATTTTCTCTG
16851 TCTTTGGTTT TCTGCAATTT GAATAATGAT ATGCCTAGGT ACAGATTTTT
16901 TTTTAATATT CATTCTGCTT GGTGTTCTCT GAGCTTCTTA GATCTGTGGT
16951 TTGGTGTCTG TCATTAATTT CAGAAAATTC CCAGTCATTA TTATTTCAAA
17001 TATTCTGTTC TCTTTCTCTT CTCCTGCTGG AAATCCAATT ATGTGTGTGT
17051 GGTGCCGTTT GAAATTGTCC CACAGCTCTT GGATATTCTG TTCATTTTTT
17101 TCACTCTTTT TTTTTTCTCT TTGCATTTCA GTTCGAGAAG TTTCTGTTGA
17151 CATTTCTTCA AGCTCATAGA TCCTTCCTCA TCTGGATCCA GTCTACTGAG
17201 GAACCATCAA AGGCATCCTC ATTTCTGTTA CAGTGTTTTC TACTTCCAGC
17251 ATTTCCTTTT GATCCCTCCT TGGAGTTTCC ATCTCTCTAC TTACATTGCC
17301 CTTCTGTTCC CCATCTGTTC TTGCCTGCTT TTCCCCTGGA GCCCTTGCCA
17351 TATTAATCAC AGTTATTTCA AATTTCCTGT TTGATAATTC CAACATGGGT
17401 GCCATATCTG GATCTGTTCT AATGTGTGCA TTGTCTCTTC ACACTCTATT
17451 TTTTTCCTTA AGACCTTGTA GTCTGCCTTG TAATGTTTGT TGAAAGCTGG
17501 ACGTGATGTA TCTGGTAATA AGAGCTGAAG TAGATGGGCC TTTAGTGTGA
17551 GGTTTATGTA ATCTGGCCAG GTTTGGGGCA GGGTTTAAAG TCTGCTGTAG
17601 CTGTGGGTAC CAGAGGCTTC ACATTTGTTT TCATTTCCCG GGTTGCCCTT
17651 GGGCTTGCCT AAATCCTCCT CCTCAGAGAG AGTCTGCGTC TTGTGGCCCT
17701 CTCTGCTGGA ATCCCTGTCA CACTGCGGAG GCCCTGTGGG TGTGTTGGGA
17751 AGATGGTGGG GAGGGAACTG TTCCACAGTC TGTGACCAAA TCTCAGTCTT
17801 GGGGGCCTGT GCCCCTTCAC AGTTGTTGAT CTGCTTTTCC CCTCCCCTTA
17851 GGTGAGACAC GCTAAAGCGG GGTACAGGCT GAAAAACAGT CCTCCCCCAA
17901 GTGAGATAAG CCTTTCCTTT GGAGAGCAAA TTCCATTTGC TGTGGAGAAT
17951 GCTCTGGGTG TATTTCACAG TGGTGACTGT CCCCATCCCA TGCCAGAGCC
18001 AGGAAGGGAT CATCCTTGGC TTCATTAAGA CCTGGCAGGG TTCCTGGAGG
18051 GGAAACCCAC AAAACGTTGG GGGCCTCATA AGACCGCAGC CGCAGGAGTT
18101 CTGCACACGG CCCCGGCCGC TCCCCAGAGC TGCCCAGGAG GTGTTCCCGC
18151 ACACCATCAG TTCTGCTCCA GGCCAGCAGG TCTCAGCTGT GACTTTGCTC
18201 ACCTGTCTCT CCAGACTTGG GGGTCGCCGC AGCGCTTTGA CCTCAGCTCT
18251 CTGATGGGTC CAGGAAAAGT AATTGATTTT CATTAGTTCA GCATTTTTCC
18301 TGTCGCAGGG ATGGGAGTGA GGCCTTCCAC GTTGTTTGCC GCAGAGCAGA
18351 AGCCAGAAGT CCTTGCTCGT CTCTGTGTTT CTTCGTCGGA GTCATGTCTT
18401 TTATTTCCTG AACGCAGGTG TGTGGTCAGA CAGGAGATTG GCAGGTATTT
18451 CCTCCTGTTC CATGGGCTAT CTTTTCTCTT TCTTGAGGGC ATCATTTGCA
```

FIGURE 3E

```
18501 GCAGGGAAGT CTTGGACCTT GACGTTGTCC GATGTGGCTG TTTCCTCCTT
18551 GGCTGCTCTG CTTTGTCTAA GGATCCATCA CCTCAGCTGA GGCCACAGGG
18601 ATTTACTCCC ACACTTTCTT CTAAGTCTTG TATAGTTTCA GCTCTTGCGT
18651 TTAGCTGTGT GATTCATTTT GGGTTAATTT TTATGAACAA TGTCAGGTGA
18701 GGGTCCAGCT TCATTCTCTC ATCGGTGGAT ATCTGACTGT CCTAGCACCA
18751 TTTGCTGAAG AGAGGATTCT TTCCCCATTG AATTGCTTTT GACATCATAC
18801 CTTGTTTTTT GACTTGCCGT TTTATCCCAT TGGTCCAGAC GTCTGCCCTG
18851 CGCCGGGACT GCACCATCTT GATAACTGTA GCTTTGTAGC AAGCTTTCAA
18901 ATCAGGATCT GTGAATCCCC CAGTTTTGTT CTTTTTGGAC ATTATTTGAC
18951 TATTCTGGGT CCCTGGCATT TCCACTGAAT GCTGAGGGGT CTGACAGTTG
19001 CATCTGAGCT GCCAAGCAGG TTTGTGGCGG TGCTAGGGAC TGAAGCCTGC
19051 TCCATTTCCC AGGCCCCTCC TCGCTGTGGG TGACATCTGG GGTCCGAGGC
19101 TGTGTCTCAG CATGTGTGAA GGTGCCACGG GTGCCCTGAG ATGGGGATTC
19151 CTGGTCCAGT TACTCAGAAA GTGCATCCAG GAGAGACCCC CGCCCTTCTC
19201 GGGATGGGAG ATGCCAGCAG AGCTTGGCTT TCAAGCAGAA ATCTGGAAAC
19251 CCTGTGGGGA GTGGCTTCGG ACTTCAGGGG ACCTGGAGCG TCACTTGTGG
19301 TTCAAGAGGT CCCTGCCCTG AGGGAGCTGG GCTCTCAAAC ACCCACACCA
19351 GTCAGTCATG GGCGAAGGGC CCCTCTGTGG CCTTCTGACT GTGTGTGTGC
19401 TGGCAAAGGG TTCCAGCCAC CCAAGGAGGA GGCAGGGGCT GTCAGAGGAA
19451 GAGCAGGGCA GATCCAGGCC AGGCACAGAC CCTGCCATGG GGTACTGCTG
19501 GCGGCGCTAC ACACACCTTG GGAAGGGAGT CCCTGTGGAA AGGGGTCGTG
19551 GTCACACATC TAGGTGACAC AGCCCGGCTT GGGCGCTGCT CAGAGCCACC
19601 CCTTCCAGAT GGTTCTGGAG CAGCTCCTCA TGCCTGGCTT GGCCTCTCTG
19651 CCTAGGAAAA CATGGCTGTG GACGTTGCAG GATGACCAAC AGCCCCTGCC
19701 ACTGGGCTGC ACACAGGGCC ACGACGGGCG CTCATGTTCT ACATCACTGG
19751 CGCCCACCCC AGCCCCTCCC ACCTTGTGTC GCTGTGAATC GCAGGATCCC
19801 AGCGGCTCAG TCGGACCCTC ATTCCTGAGT AGTCTGAGCC TTAGGTCACC
19851 GTCACCTTCT CAGGCCGGCC CGAGTTTGCA GACTTGTCTG TCTATATCAG
19901 GGTTAGACCA GAGAGTGCTG AGACACAGCA GATCACCCAG CCTGTCCTCT
19951 TCTTGATGAC TAAGGACAGG TCCCCTGCCA GGATCGTGAC TCCTTTAGGG
20001 GAGGCCACAG TGACAGGGCA AAGCCTGGAG GGAGAGAGCC ACATGGAGAG
20051 GAGAGGGCTG CCCGCAGAGA GCGTGGGAGT CTGCCGGCTT CTTCTCGAGT
20101 CCTTGGCAAG GTGCTGGCCG CTCACACCGT GTACGTGTGG GGAACGCCCA
20151 GGACCAGGGT GACACCACCA GGAGGAGCGG GCGGGACAGT CCCCACTCAG
20201 GGCTAGGAAG AGAGAGTCCA CATTCCCGCT GCCAGGGTGA AACCCTCACA
20251 CCACCACAGA TCCAGGAGAA ACACGGAGGG CACTGCCTCG GGGTGGGGAA
20301 CGTGAGCTGC TCCCCAAACC CAAGAAATGT GTTGAGCCCT ATGCTTCCTT
20351 CCTCGTGGGA AGAGGCGCAA GGTGAGACCG CTTGTCCTTT ATCTTGGCGG
20401 GCACGTCCGA GTGTGACGCA CGTCAGCAAA TCCCTGAACC GTTCATCGGA
20451 GAACAGCCTT CTGCATCTCC CACACTCTGT TCGTGGGTTT ACAGGGTGTC
20501 CAGAGTACTT GCCACTTGGC AATCAGCATT AATAGGATCC ACAGGGCCAG
20551 GCATGGTGGC TCACGCCTGT AATCTCAGCA CTTTGGGAGG CCGAAGCAGG
20601 CGGAGCACCT GAGGTCAGGA GTTCAAGACT GGCGTGGCCA ACATGGTGAA
20651 ACCCCGTCTC TACAAAAATT AGCTGGGTAT GGTGGTGGGC ACCTGTAATC
20701 CCAGCTACTT GGGAGGCTGA GGCAGCAGAA TCACTTGAAC CTGGGAGGCG
20751 GAGTTGGCAG TGAGCCAAGA TCACGCCACT GCACTCTAGT CTGGGTGACA
20801 GAGCGAGACC AAAAATAGGA TCCATGGATA GCAGGCAAGA GTGTCCAGGT
20851 GTTCGAGGCA CAGACGACAC TGTGACAGGG AAGAGTCCCC TTAGCCCTGG
20901 CTGGGGCCGT GAAAGCATGC TGTTGTCCGT CTCGGGTGAA CGCAGACTGT
20951 TGTCACGCAT TGCATAACGA TGCTTCCGTC ACTGGCCAAT CGCATGGGGG
21001 GGTGGTCCCG TAAGATGGTA ACACTGGGTT TTGCTGTACG TTTTGTATGT
21051 CTAGATAGGG TTGAGCGTTC TGGTGTGTAC CCACTCACAC GTCCCCTCCG
21101 ACCTTCAGAG CCCAGCTCCC TCCCTCCCAG GGCCTTGGCT GTGACGTGGG
21151 TGACTTCCTA TGGATCTGAG GTTCTGTGGT CCTCACAAGT GGGCATCCTC
21201 TGGCCTCAGC TGCAGGAGGT GGGGGCCCTT TTAATGCCAC CCGAGGCCTG
21251 CGACTCCCTG CACTTTTCAC TGTGACTTGG CTCATCTGGG TCTGTCATTT
21301 GCTCACGCGT TGGTAGTGAC CAACGTCACC ATCCAAGTTC ACGGTCACCA
21351 TAATGATGCT TTCCCCACAC CATGCCAGCG CTGAGCGGCC AGCACCCCCT
21401 CCACCCCACC CACGCCCTCC CCCGACCCCT GCGTCCTGGG AAGTGGTCCT
21451 GCTGCCTGAA GGAAGTAGTC CTGCCTGCCC ATCACACACC AGTAAGGGTG
21501 GGTCCTGCCA GGGGCAGCCT CCGTCCACAA GCTTGCCCTG AGGACCTGCT
21551 TCTAAGACAG CCCTGGTTCC AGGATTCTCT GGGCAGGGCC CCAGAAGCAG
21601 GCCTGGGACA GGTGTGTGTG TGCTGTGATG AGGGGCTGGG AGAACCCGGT
21651 ATGAGACGGG AAAGGCCCGG CAAGGGAGTG GTTTCCAGCA AAGTCCCGCA
21701 GAGAACAGCT TCTGCCTGGT CCTGCAGGCC CCACGGAGCA AGTCCAAGCC
21751 ATCCACCCAG AGGCAAGGGA GCTGGGCCTT GGCATCCTCG GGCTTGGGTG
21801 AGTCACCCGC AGAGATGCGA GCTCCCCGGG CAGTCTGGCT GCTGGAGGGC
21851 CGGGGCACTT CAAATAGCCC AGAGGCCGTC ATCCAAAGCC ACAGGTGGAG
21901 GCCCGATGGG GATGCCCAGA CACTCACTTG AGGGGACATG GGCGGAACCT
21951 GGACAGCGTC CCCCACGCTC ACGTGTGCCT TTCCATCCAC AGGAGAATGA
22001 CTATCTTCAA GACTGCCTCG ATGCCATCCA GCAAGACTTC GTGATTTTTA
22051 ACAGAGAAAA GTGAGTGTGT TGGGGGTGGG TGGGCGTGGT GGCAGAGAGG
22101 AGGAAAATGG GGCTAAGGTT AAGGTTTTCT TGGCCACGTG AGCGGGCACC
22151 TGTGGGCCTG GGGTGCGTGG CCCTGCTCTC TTTGGGGACT CTGAGCAGCA
```

FIGURE 3F

```
22201 GCTATGGAGG GGAGCGGCGG GAGGCCCCTG CCAGGCTCTG GCATGTTTGT
22251 GCTCCACGCG GGGCCCGTGG CTGGAATCTT CTGGGGAGAG ACACATCATT
22301 TGCCCAGATG AGGGGTGGTG ACTTCCTAGG AGGCCCCATC AGAGCCACGT
22351 CAACTCCCCC ACCCAGGCAC GCCCTCAGTC TCTCAGCAGA CCTTTCCTGA
22401 ATGTCAGGCC CCAGGGGACA GAAAGGGCAC AGATGACTGG CAGCAGGCAA
22451 GGCAGGCCAG AAATAGCAGC AGCTGCCACG GTGGGGCCCA AGGGAGGATG
22501 GATGCTCCCT CTGCCCGCAC GGGGCAAGGA GGGGCCTCCTG GAGGAGGTGG
22551 GTCTGAGCTC TTATGGACAG GACGTGCAGG GCAGCACGTG CAGACGGCTG
22601 AGGGCACTGA CTGGCACCCT GGGGATCAGA CGACCGGGTG AAGAATGAGG
22651 CTTAGCCGAG CCTCATTCCC AAGTCACTGA CCTATGGCAC CTGCACAGTC
22701 AGGCCTTTCG GCTTCTGGCT GGAAACATGC GGCCTCCTCGC CAGCATGCTC
22751 ACGTGCCCCC ACCCGTCCCC AGGCTCCCTG CCAGTGTGTC GGGAGCATGG
22801 CCTCTCCAGC AGACACCGAG CCTGTGGCCC ACGTTTGGGC ATCCACGCCA
22851 TGGCCTATCC CATGAGCCCG TGGGCAGGTC ATGGGACCGT GAGGCCAGGG
22901 AGGTGGGGGC ATAACGCCCT CCATGTGTTC CTGCCACCCC AGGCTGAAGA
22951 GGAGCCAGGA CCTCCCGAGG GAGCCTCTCC CCGCCCCTGA GTCCAGGGAT
23001 GCTGCGGAGC CTGTGGAGGA CGAGGCGGAA CGCTCCGCCC TGCCCATGTG
23051 CGGCCCCATT TGCCCCTCGG CCGGGAGCGG CTAGGCCGGG ACGCCCGTGG
23101 TCCTCACCCC TTGAGCTGCT TTGGAGACTC GGCTGCCAGA GGGAGGGCCA
23151 TGGGCCGAGG CCTGGCATTC ACGTTCCCAC CCAGCCTGGC TGGCGGTGCC .
23201 CACAGTGCCC CGGACACATT TCACACCTCA GGCTCGTGGT GGTGCAGGGG .
23251 ACAAGAGGCT GTGGGTGCAG GGGACACCTG TGGAGGGCAT TTCCCGTGGG
23301 CCCCCGAGAC CCGCCTAGAT GGAGGAAGCG CTGCTGGGCG CCCTCTTACC
23351 GCTCACGGGG AGCTGGGGCC ATGGATGGGA CAGGAGTCTT TGTCCCTGCT
23401 CAGCCCGGAG GCTGTGCACG GCCCTCGTCA CAAGGTGACC CTTGCAGCAC
23451 AGGCCGCGGG TGCCCCAGGC TCGGCTCAGT TCTTGGAGGT CAAGGGCATG
23501 GGTTGGGGTA GTGGGTGGGG AGGTGAATGT TTTCTAGAGA TTCAAACTGC
23551 TCCAGCAATT TCTGTATAGT TTTCACCTCT GAGAATTACA ATGTGAGAAC
23601 CGCTCGATGT TGCATGTTCT GCGTACGTCC TGTGTCTGCC TGGCCGTCAG
23651 GCCGGTGCCT GCCGTTTCTG GTTGGCCTGG ACTTGGGGCA GCCAGTGGGG
23701 TGGGCAGCTC CTCAGGGCAG AGCTCCCGGA CCATGGCTTT GGGGTGGGTG
23751 CCTGTCTCCG TGGCCCTGGA GCCGTAAGGC TGTGGAAGGC AGAGACGGTC
23801 CTGGAGGCAG AGGAGCCCAG GACAGCACCG TGCACCGTGG AGCCGCCGCA
23851 GTGCCGGGCA GTGCTTGGCC CTCCATAAAG GGACGTATCC CTCTCACTGT
23901 GGCTGGGTGG TTCTGTGGTT GGAACTGTAA CTAACTGGGT AAACGGCCTG
23951 TGTGCTTCTC TCTGGTCTCG CTGGAGGAGG ACGGGCTCAG CCCGTCAGCC
24001 CAGCGCTCCA GACAGGCCTG TGCTGGTTTC CTCTGAGGAA ATGGGTGTGG
24051 CGGGTCTGTG CCCCTTCCCA GGACAGCGGC CATAGTGGAC ATGTGCCTAG
24101 ACCTGTGTCC ATGAGCCCCA CTGCACCCCT GGCAAACAGG GCCCTCCCGT
24151 CCTTGGCTGG GCTGCGAGAT GGAGATGACA ACGGCCAAAG AACATTTGGG
24201 GAAGAACCGG CCATGCCACG AGCAGAGTCA GAAGTCCGAG GGGATAGAAT
24251 GCAGCTTCCC GTCCCCCACC CACCCCTGTC CTCTGAATCA TGGCAGAAAC
24301 TAGCCTTCCA GCCCTCAGCA GCTCACATGG GGGACACGGC ACCCAAATCA
24351 CCACCAGGAA GGGTGGCCCG GTCTCTGCGA GGGCCCAGAG GCGCCGTGTA
24401 CTCGGTGGAA GTCTGGCGAT GTCAGAGACA GGCTCGGGGC AAGGACAGGT
24451 GTGGGGGTTT GAATAAGTGC ATTTGGGGAA CATGGCAGGG TGGTGCACCT
24501 TGTCCTTCTT CGAGACACTG GTGAGGTGTG GGTGCTGTCT GGTTCCCTTG
24551 ATCGCCCCCC ACACTGGGGC AGAGTGGGAG ATGCTGGTGT GGGGGACATC
24601 AGCTCCCACA TCTGGGCCAG AGGGAGCCCC GGGAAGGAAA TGCTGAGGGC
24651 CCAGGGCCTT CGCCTGGGAT CTGCACAGCT TTATAAGCAG CCCAGGGTGA
24701 GAGATGGGCC TGTCTGTGGT TCCCAGAGAC CACGGCAGGA AATTCTCTGT
24751 CACCATCGGT GCATGGGCAG GGGCCAGAGA ACCGGTGGCA CAAGGTGTCC
24801 CTGGCTCTCT GCTCAACAAA CAGCGAGTGC CCAGTGACTG CGACCAGGCC
24851 CCGCTCTTGG GATGAGGACA ACCGTCTGGG AACGTCCACG CACCCTTTAT
24901 GAGACACAGC ACGCGCCAGC ACCGCAGTCA CACACCGGGG GCTCGGGTCA
24951 GCCTCATAGC TGCCCGGCCT TGAGTGCTGG GCCTGCGTCT GTGAGCAGCG
25001 CCCACCTGGG TGGCGGTGGT GTGTGCTTCA CTTCCACACA CTCCCGTGCA
25051 TGCTCCCCGG CCTTCTGGGG TGGCTCGGGC TGTCCGGTGA GAATGTAGGC
25101 GGGGGGGGGG GGGGCCTCT GTCCCGCCCT GGATGTTGGC TGCCCTCTGC
25151 CCCGCCCTGG ATGTTGGCTG CCCTGCCGGC CATCTTCCCT GTGAGAGGGT
25201 GCGCTCCTCC CTGCCATTGA GGGGAGAAGA GCTGCGGCTG CAGGAGTCGT
25251 CGGACCAGCC CACAGCCAGC CAGGCCCCGC CTGTGCAGAG ACGGCGTGGG
25301 GGAGAGGAGA CGGGGCCTTC CTTCCATGCA CAGGCGGCTT CAAACCCAGA
25351 CGTCTTTAAT GGGCCTGATT CACATCGAG GCAGGATGAC TGCCTGTCCA
25401 GGCGGTGGGT GGCATGCACA GGTTCCTGGC TACAGTGTCC TCAGTGTACA
25451 AAGCTGCTAC TAAGAAGCCT ACGGAAATAC ACAATCTGTA ATAAGAGGAC
25501 AGTGTCTTCC TAAAGGATCG CAAAACTTCC CTGGATGAGG GCTACATGGA
25551 AGCTTAGGTG TGGGCCTTGG GGTGCGTAAA AGGGACCCTC CACGGGCGGG
25601 GCT
```

(SEQ ID NO: 3)

FEATURES:
Start: 3000
Exon: 3000-3118

FIGURE 3G

```
Intron:   3119-4126
Exon:     4127-4300
Intron:   4301-5310
Exon:     5311-5348
Intron:   5349-5478
Exon:     5479-5568
Intron:   5569-5758
Exon:     5759-5862
Intron:   5863-6571
Exon:     6572-6688
Intron:   6689-8108
Exon:     8109-8234
Intron:   8235-8322
Exon:     8323-8454
Intron:   8455-21992
Exon:     21993-22060
Intron:   22061-22942
Exon:     22943-23081
Stop:     23082
```

SNPs:

DNA Position	Major	Minor	Domain	Protein Position	Major	Minor
1931	G	A	Beyond ORF(5')			
4232	G	A	Exon	75	T	T
4309	G	T	Intron			
4622	G	A	Intron			
4708	T	C	Intron			
5787	C	T	Exon	150	N	N
5884	C	T	Intron			
8481	C	T	Intron			
8754	G	A	Intron			
8847	G	A	Intron			
11114	G	A	Intron			
19741	A	G	Intron			
20908	C	T	Intron			
22728	T	C	Intron			
23406	C	T	Beyond ORF(3')			
24078	G	A	Beyond ORF(3')			
24777	G	A	Beyond ORF(3')			

Context:

DNA
Position

```
1931      AGTTCAGTGGTACAGTCGTGGCTCACTGCAGCCTCAACCTCTCAGGCTCAAGCAATCCTC
          CCTTCTCAGTCTCCTGAGTAGCTGGGACTACGGGTGTGTGCCACCACGCTTTCCGGGCAT
          GCAGCCAGGAGCCCAGGGCCATCTGTGGCCCACCTTGAGATCCAGAATCATCCATTTCCT
          CCAGGCCCCCTGCTGGGCTCCAACTCCTTGAGGACCAGAGAGCAGAGGTTGTGGAAGGCC
          TTGGAAACGGGTCTGGATTACCTGTCCTGGGAAGGTCTCTCCCAACCTGAGTGTCAGACA
          [G,A]
          GGGTTAGCTCTGCTGCTCACAATTTTTGTGCCTTAATTCCTGGCTTCCCTTTGGGGATCTT
          CATCCTCAATTCTGATTGACATCCTTGGCCACAAGGGACCCCCCTGCTCATTGATGCTTC
          TCACCCGTCACCTCACTCTCATCCTCACTGCTAAGCAATTAGCCGTGTGTTTGCGGCATC
          AGTGTTGACACCGATGATCCATGCTCAGAGGGTACAGGCCTGAAGAGCTATGTGGGGACT
          GGCGCCCCGGAGGGGGTCCCGCTGTGGTGGCAGCGGTGGCCCCCAAGCCCCACGCTCACT
```

```
4232      CCTCAGGATCTGGGCTCAGAGCTGCTCTGGGCTGGGGTGCAATCAGTGCCTTGGGCAGGC
          CCCTCCTCCTGGGAATGCCTGGTGGCTGATGCTGGGGTGGGGCTGTGGTCCTTAGGGGGA
          GTGTGTCAGCTGTGGGAGCAGCCATGACTGGCTCCCCAGCTGTGCGCACAACAGGCCTTC
          CATCGGTGCCCACAGGTACTCCTTCCAGGACGAGGAGGACATGTTCATGGTCGTGGACCT
          GCTACTGGGCGGGGACCTGCGCTACCACCTGCAGCAGAACGTGCAGTTCTCCGAGGACAC
          [G,A]
          GTGAGGCTGTACATCTGCGAGATGGCACTGGCTCTGGACTACCTGCGCGGCCAGCACATC
          ATCCACAGGTGTGTGCGTGGCAGACGGCGCAGGTACCTGCTGAGGTGGGCGGGGCTGAAG
          CAGCCTTAGGTCAGGCTGCCGGCACGGCCGGCCGTACTCCCTCAGAGCGGGTCTAGCTCCT
          CTGCCCCACCCTTGCCTGAGTGCCTGCCCCCAGCTGTGGCACCTGTGCCGACCAGGTCAG
          CCCCATAGCTGTGTGCCTGGTGTCCATCTGGGGGGGACCTCGTCCCAGCAGCCCCAGCTGA
```

FIGURE 3H

4309 CCTGGTGGCTGATGCTGGGGTGGGGCTGTGGTCCTTAGGGGGAGTGTGTCAGCTGTGGGA
GCAGCCATGACTGGCTCCCCAGCTGTGCGCACAACAGGCCTTCCATCGGTGCCCACAGGT
ACTCCTTCCAGGACGAGGAGGACATGTTCATGGTCGTGGACCTGCTACTGGGCGGGGACC
TGCGCTACCACCTGCAGCAGAACGTGCAGTTCTCCGAGGACACGGTGAGGCTGTACATCT
GCGAGATGGCACTGGCTCTGGACTACCTGCGCGGCCAGCACATCATCCACAGGTGTGTGC
[G,T]
TGGCAGACGGCGCAGGTACCTGCTGAGGTGGGCGGGGCTGAAGCAGCCTTAGGTCAGGCT
GCCGGCACGGCGGCCGTACTCCCTCAGAGCGGGTCTAGCTCCTCTGCCCCACCCTTGCCT
GAGTGCCTGCCCCCAGCTGTGGCACCTGTGCCGACCAGGTCAGCCCCATAGCTGTGTGCC
TGGTGTCCATCTGGGGGGGACCTCGTCCCAGCAGCCCCAGCTGAGACTGGGCACAGTGGGC
TGTTAGCCCTGGTGGACAGACCACCAGGCTGGGTCACAGCAGGTGGCCTTCACCTGGTCC

4622 CAGGTACCTGCTGAGGTGGGCGGGGCTGAAGCAGCCTTAGGTCAGGCTGCCGGCACGGCG
GCCGTACTCCCTCAGAGCGGGTCTAGCTCCTCTGCCCCACCCTTGCCTGAGTGCCTGCCC
CCAGCTGTGGCACCTGTGCCGACCAGGTCAGCCCCATAGCTGTGTGCCTGGTGTCCATCT
GGGGGGGACCTCGTCCCAGCAGCCCCAGCTGAGACTGGGCACAGTGGGCTGTTAGCCCTGG
TGGACAGACCACCAGGCTGGGTCACAGCAGGTGGCCTTCACCTGGTCCATTTAACTGAAG
[G,A]
CTCCTGTTTGCCCATCCACCACATCCCAGGGAATCCAAACTAATTTTAACATTAGCTTAA
AGCAGATGAAATTAGGAAGCAGAGCTGGTGTGATGGCTCTGAAAATAAAATTTAAAAAAA
GAAAATAGGAAGCAGATTATGAAGGAAGTGAAATTGGGAAGCAGAAATTAGGCTGAAATT
CCGCAGCAATGGAACAAAATGAAAATATCTGTGAGGTATATTTTAAAGTCGAATGGACTG
GTGTTTGCATTTCTGCTCTTGGGGACTCGGATGTCTGATTATGACCTAGGCACCAGTCAC

4708 CTCCTCTGCCCCACCCTTGCCTGAGTGCCTGCCCCCAGCTGTGGCACCTGTGCCGACCAG
GTCAGCCCCATAGCTGTGTGCCTGGTGTCCATCTGGGGGGGACCTCGTCCCAGCAGCCCCA
GCTGAGACTGGGCACAGTGGGCTGTTAGCCCTGGTGGACAGACCACCAGGCTGTTAGCCCTGGGGACAGT
GCAGGTGGCCTTCACCTGGTCCATTTAACTGAAGACTCCTGTTTGCCCATCCACCACATC
CCAGGGAATCCAAACTAATTTTAACATTAGCTTAAAGCAGATGAAATTAGGAAGCAGAGC
[T,C]
GGTGTGATGGCTCTGAAAATAAAATTTAAAAAAAAGAAAATAGGAAGCAGATTATGAAGGA
AGTGAAATTGGGAAGCAGAAATTAGGCTGAAATTCCGCAGCAATGGAACAAAATGAAAAT
ATCTGTGAGGTATATTTTAAAGTCGAATGGACTGGTGTTTGCATTTCTGCTCTTGGGGAC
TCGGATGTCTGATTATGACCTAGGCACCAGTCACTGAGCACTGGCTGTGTACCTGGAAAA
GTTGGGACAAAGCAAGAGCCGAGGTGGCTTGGTCTCCTAGAGGCCGAGTCTTGGAGGGGG

5787 CACCTGACCGACTTCAACATTGCCACCATCATCAAGGACGGGGAGCGGGCGACGGCATTA
GCAGGCACCAAGCCGTACATGGGTGAGCCCGAGCTGGGGTTCCAGATGGGAGCTGGCTTC
CTCCAGGTGGGAAGGACAAGACCTCGGTGGCTTCTCTGTCCCACCCTGGAGGCAGCCTGG
TCTCGGGATGTGGCCTCAAGGTGCCGGCCCTGTGCCCACGGGTCCGGGCTGTGACCCCGT
GGCAGCTGTTTTTCCTTCTTTCTGTCGGAAAGCTCCGGAGATCTTCCACTCTTTTGTCAA
[C,T]
GGCGGGACCGGCTACTCCTTCGAGGTGGACTGGTGGTCGGTGGGGGTGATGGCCTATGAG
CTGCTGCGAGGATGGGTATGGACCCCCTGCAGCCCCCGGGCTTGGCTGCCAGGCCCCTGC
TCTCTGCCCCCACCAGTGCTGGGGAGGGGGTGGCTGCCCCAGTGCCCAGGTGCGCAGGGA
TGTCTCCACTGTGTCTGAGGAGTCACGCTTTTATCGAAGTGTGTAGTTGGTGATGGAATG
CCTGAGCAGGAGGAGGAAGGACAGACTCACTGTGGTTTCCCGGGGCCGCTGCTGGTGCCT

5884 GGTTCCAGATGGGAGCTGGCTTCCTCCAGGTGGGAAGGACAAGACCTCGGTGGCTTCTCT
GTCCCACCCTGGAGGCAGCCTGGTCTCGGGATGTGGCCTCAAGGTGCCGGCCCTGTGCCC
ACGGGTCCGGGCTGTGACCCCGTGGCAGCTGTTTTTCCTTCTTTCTGTCGGAAAGCTCCG
GAGATCTTCCACTCTTTTGTCAACGGCGGGACCGGCTACTCCTTCGAGGTGGACTGGTGG
TCGGTGGGGGTGATGGCCTATGAGCTGCTGCGAGGATGGGTATGGACCCCCTGCAGCCCC
[C,T]
GGGCTTGGCTGCCAGGCCCCTGCTCTCTGCCCCCACCAGTGCTGGGGAGGGGGTGGCTGC
CCCAGTGCCCAGGTGCGCAGGGATGTCTCCACTGTGTCTGAGGAGTCACGCTTTTATCGA
AGTGTGTAGTTGGTGATGGAATGCCTGAGCAGGAGGAGGAAGGACAGACTCACTGTGGTT
TCCCGGGGCCGCTGCTGGTGCCTGCAGGCCAGCCTCTGTGGGGGTGGACAAGGCTGAGAA
CTGGCCAGCAGGGGTGCTGCCTCGGAACTTTCCACAAAAAGTTTCTTTTGGGGCCCTGTG

8481 GTGCTGTGGGACCACCTGAGCGAGAAGAGGGTGGAGCCGGGCTTCGTGCCCAACGTAAGC
CTGTGGGCGGCTCAGGTGGGGGGGCCCTGGGGATGGATGTGGCGTCCTCCACGGGCCGGGG
CTCAGCACCCATCCCTCTGTAGAAAGGCCGTCTGCACTGCGACCCCACCTTTGAGCTGGA
GGAGATGATCCTGGAGTCCAGGCCCCTGCACAAGAAGAAGAAGCGCCTGGCCAAGAACAA
GTCCCGGGACAACAGCAGGGACAGCTCCCAGTCCGTGAGTGCCAGGGCAGGCTCAGGGCG
[C,T]
GGCGGCGGGCTGGGCTTGGGGCTCCTCTCTACCACCGAGCAAGGTGTGTGGGGACCCCTG
ACAGTGCACACGTCTCGGAAGTCCAGCAGACCGTTTCCTGAAGTCCTGAGAAGGCCAGAG
ACCTCCCTTCTGCCTTTCCCAGCCCCCACCTCGCTCCTTATGAAGCAGGTGGGCAGGGAC
AACCAGGGCTGGGGTTATGAGTGCACGGGGATGGCCATGTGAAGCCTTCGTGCTTGCCCA
GGTGTGCTGGTGTTGGTTGTGTGTGCGGGGACGGCTATGTGAAGCCCTCACACTCGCCCA

8754 CGTGAGTGCCAGGGCAGGCTCAGGGCGCGGCGGCGGGCTGGGCTTGGGGCTCCTCTCTAC
CACCGAGCAAGGTGTGTGGGGACCCCTGACAGTGCACACGTCTCGGAAGTCCAGCAGACC

FIGURE 31

```
            GTTTCCTGAAGTCCTGAGAAGGCCAGAGACCTCCCTTCTGCCTTTCCCAGCCCCCACCTC
            GCTCCTTATGAAGCAGGTGGGCAGGGACAACCAGGGCTGGGGTTATGAGTGCACGGGGAT
            GGCCATGTGAAGCCTTCGTGCTTGCCCAGGTGTGCTGGTGTTGGTTGTGTGTGCGGGGAC
            [G,A]
            GCTATGTGAAGCCCTCACACTCGCCCAGGTGCGTCGGCATCAGGTATGTGTGCCGGGACA
            GCCATGTGAAGCCCTCACACTCACCCAGGTGCGTCGGCATCAGTTGTGTGTGGGGACG
            GCCATGTGAAGCCCTCACACTCGCCCAGGTGTGCTGGCTTTGGTTGTGTGTGCAGGGATG
            GCCACATGAAGCCCTCACACTCGCCCAGGTGCGTCAGCATCAGGTGTGTGTGCGGGGACG
            GCCATGTGAAGCCCTCACACTCGCCCAGGTGCGTTGATGTTGTGTGTGCAGGGATGGCCA

    8847    GCACACGTCTCGGAAGTCCAGCAGACCGTTTCCTGAAGTCCTGAGAAGGCCAGAGACCTC
            CCTTCTGCCTTTCCCAGCCCCCACCTCGCTCCTTATGAAGCAGGTGGGCAGGGACAACCA
            GGGCTGGGGTTATGAGTGCACGGGGATGGCCATGTGAAGCCTTCGTGCTTGCCCAGGTGT
            GCTGGTGTTGGTTGTGTGTGCGGGGACGGCTATGTGAAGCCCTCACACTCGCCCAGGTGC
            GTCGGCATCAGGTATGTGTGCCGGGACAGCCATGTGAAGCCCTCACACTCACCCAGGTGC
            [G,A]
            TCGGCATCAGTTGTGTGTGTGGGGACGGCCATGTGAAGCCCTCACACTCGCCCAGGTGTG
            CTGGCTTTGGTTGTGTGTGCAGGGATGGCCACATGAAGCCCTCACACTCGCCCAGGTGCG
            TCAGCATCAGGTGTGTGTGCGGGGACGGCCATGTGAAGCCCTCACACTCGCCCAGGTGCG
            TTGATGTTGTGTGTGCAGGGATGGCCATGTGAAGCCCTCACACTCACCCAGGTGCGTTGA
            TGTCAGTTGTGTGTGCAGGGACAGCCATGTGAAGCCCTCAGACTAGCCCAGGTGTGTCGG

    11114   AGACGCATCCCCCCAGCCCCTGCTTCGCTGCCGCGTGGGGTCGGCCTGTGTGTGCGTCTC
            CATCTCCTCCCCTCTTCTCATAAGGACACCAGGCATTGGATTTAGGGCCCACCCTGATCC
            AGTATGGCCCCATCTTATCTTGATGATATCTGCAAAGACCTCACTTCCAAATGAGGTCAC
            ATTCACAGGTACCCAGGATTAGAATTTGAGTGTGTCATTTTTGGGGACACAGTTTGGCCC
            ATACCACCAGGATGTGGCTGATATTCACCAAGGAGTAGCTATGGTTGTGTGTGTTGATGTCA
            [G,A]
            GGTGACGGTGATGACCCTGGGTCCCTCGGTGGTCCCCTTGCCCCTGAGTCTGCCTGAGCC
            TGTGGTGGATGTCCTGGGAAACTCTTGTGCCTCAGCCCCCGTGCAGCCTCCTCAGACCTG
            GTGGGCCCTGTGTTGCTCCTGGGCAGAAGACGGGTGTCAGTCCCCTCCTCACCATGATGT
            GGGGGGCAGGGGTGGGGTCATGCCCTGGGTGCCCTGATTTTGGGGGGAACACGGCCCCCC
            AGTGGGTCAGGCTCCCATCCTCGCCCCTCCTCCAGGACGGCTGCCGGCAGCCCTGGGTGT

    19741   GTCAGAGGAAGAGCAGGGCAGATCCAGGCCAGGCACAGACCCTGCCATGGGGTACTGCTG
            GCGGCGCTACACACACCTTGGGAAGGGAGTCCCTGTGGAAAGGGGTCGTGGTCACACATC
            TAGGTGACACAGCCCGGCTTGGGCGCTGCTCAGAGCCACCCCTTCCAGATGGTTCTGGAG
            CAGCTCCTCAGGCTTCTGGTGGCCTCTCTGCCTAGGAAAACATGGCTGTGGACGTTGCAG
            GATGACCAACAGCCCCTGCCACTGGGCTGCACACAGGGCCACGACGGGCGCTCATGTTCT
            [A,G]
            CATCACTGGCGCCCACCCCAGCCCCTCCCACCTTGTGTCGCTGTGAATCGCAGGATCCCA
            GCGGCTCAGTCGGACCCTCATTCCTGAGTAGTCTGAGCCTTAGGTCACCGTCACCTTCTC
            AGGCCGGCCCGAGTTTGCAGACTTGTCTGTCTATATCAGGGTTAGACCAGAGAGTGCTGA
            GACACAGCAGATCACCCAGCCTGTCCTCTTCTTGATGACTAAGGACAGGTCCCCTGCCAG
            GATCGTGACTCCTTTAGGGGAGGCCACAGTGACAGGGCAAAGCCTGGAGGGAGAGAGCCA

    20908   CCTGAGGTCAGGAGTTCAAGACTGGCGTGGCCAACATGGTGAAACCCCGTCTCTACAAAA
            ATTAGCTGGGTATGGTGGTGGGCACCTGTAATCCCAGCTACTTGGGAGGCTGAGGCAGCA
            GAATCACTTGAACCTGGGAGGCGGAGTTGGCAGTGAGCCAAGATCACGCCACTGCACTCT
            AGTCTGGGTGACAGAGCGAGACCAAAAATAGGATCCATGGATAGCAGGCAAGAGTGTCCA
            GGTGTTCGAGGCACAGACGACACTGTGACAGGGAAGAGTCCCCTTAGCCCTGGCTGGGGC
            [C,T]
            GTGAAAGCATGCTGTTGTCCGTCTCGGGTGAACGCAGACTGTTGTCACGCATTGCATAAC
            GATGCTTCCGTCACTGGCCAATCGCATGGGGGGGTGGTCCCGTAAGATGGTAACACTGGG
            TTTTGCTGTACGTTTTGTATGTCTAGATAGGGTTGAGCGTTCTGGTGTGTACCCACTCAC
            ACGTCCCCTCCGACCTTCAGAGCCCAGCTCCCTCCCTCCCAGGGCCTTGGCTGTGACGTG
            GGTGACTTCCTATGGATCTGAGGTTCTGTGGTCCTCACAAGTGGGCATCCTCTGGCCTCA

    22728   CACAGATGACTGGCAGCAGGCAAGGCAGGCCAGAAATAGCAGCAGCTGCCACGGTGGGGC
            CCAAGGGAGGATGGATGCTCCCTCTGCCCCGCACGGGGCAAGGAGGGCCTCCTGGAGGAGG
            TGGGTCTGAGCTCTTATGGACAGGACGTGCAGGGCAGCACGTGCAGACGGCTGAGGGCAC
            TGACTGGCACCCTGGGGATCAGACGACCGGGTGAAGAATGAGGCTTAGCCGAGCCTCATT
            CCCAAGTCACTGACCTATGGCACCTGCACAGTCAGGCCTTTCGGCTTCTGGCTGGAAACA
            [T,C]
            GCCGAGCCTCGCCAGCATGCTCACGTGCCCCCACCCGTCCCCAGGCTCCCTGCCAGTGTG
            TCGGGAGCATGGCCTCTCCAGCAGACACCGAGCCTGTGGCCCACGTTTGGGCATCCACGC
            CATGGCCTATCCCATGAGCCCGTGGGCAGGTCATGGGACCGTGAGGCCAGGGAGGTGGGG
            GCATAACGCCCTCCATGTGTTCCTGCCACCCCAGGCTGAAGAGGAGCCAGGACCTCCCGA
            GGGAGCCTCTCCCCGCCCCTGAGTCCAGGGATGCTGCGGAGCCTGTGGAGGACGAGGCGG

    23406   ACCCCTTGAGCTGCTTTGGAGACTCGGCTGCCAGAGGGAGGGCCATGGGCCGAGGCCTGG
            CATTCACGTTCCCACCCAGCCTGGCTGGCGGTGCCCACAGTGCCCCGGACACATTTCACA
            CCTCAGGCTCGTGGTGGTGCAGGGGACAAGAGGCTGTGGGTGCAGGGGACACCTGTGGAG
            GGCATTTCCCGTGGGCCCCCGAGACCCGCCTAGATGGAGGAAGCGCTGCTGGGCGCCCTC
```

FIGURE 3J

```
         TTACCGCTCACGGGGAGCTGGGGCCATGGATGGGACAGGAGTCTTTGTCCCTGCTCAGCC
         [C,T]
         GGAGGCTGTGCACGGCCCTCGTCACAAGGTGACCCTTGCAGCACAGGCCGCGGGTGCCCC
         AGGCTCGGCTCAGTTCTTGGAGGTCAAGGGCATGGGTTGGGGTAGTGGGTGGGGAGGTGA
         ATGTTTTCTAGAGATTCAAACTGCTCCAGCAATTTCTGTATAGTTTTCACCTCTGAGAAT
         TACAATGTGAGAACCGCTCGATGTTGCATGTTCTGCGTACGTCCTGTGTCTGCCTGGCCG
         TCAGGCCGGTGCCTGCCGTTTCTGGTTGGCCTGGACTTGGGGCAGCCAGTGGGGTGGGCA

24078    GGCTGTGGAAGGCAGAGACGGTCCTGGAGGCAGAGGAGCCCAGGACAGCACCGTGCACCG
         TGGAGCCGCCGCAGTGCCGGGCAGTGCTTGGCCCTCCATAAAGGGACGTATCCCTCTCAC
         TGTGGCTGGGTGGTTCTGTGGTTGGAACTGTAACTAACTGGGTAAACGGCCTGTGTGCTT
         CTCTCTGGTCTCGCTGGAGGAGGACGGGCTCAGCCCGTCAGCCCAGCGCTCCAGACAGGC
         CTGTGCTGGTTTCCTCTGAGGAAATGGGTGTGGCGGGTCTGTGCCCCTTCCCAGGACAGC
         [G,A]
         GCCATAGTGGACATGTGCCTAGACCTGTGTGTCCATGAGCCCCACTGCACCCCTGGCAAACA
         GGGCCCTCCCGTCCTTGGCTGGGCTGCGAGATGGAGATGACAACGGCCAAAGAACATTTG
         GGGAAGAACCGGCCATGCCACGAGCAGAGTCAGAAGTCCGAGGGGATAGAATGCAGCTTC
         CCGTCCCCCACCCACCCCTGTCCTCTGAATCATGGCAGAAACTAGCCTTCCAGCCCTCAG
         CAGCTCACATGGGGGACACGGCACCCAAATCACCACCAGGAAGGGTGGCCCGGTCTCTGC

24777    GGAACATGGCAGGGTGGTGCACCTTGTCCTTCTTCGAGACACTGGTGAGGTGTGGGTGCT
         GTCTGGTTCCCTTGATCGCCCCCCACACTGGGGCAGAGTGGGAGATGCTGGTGTGGGGGA
         CATCAGCTCCCACATCTGGGCCAGAGGGAGGCCCCGGGAAGGAAATGCTGAGGGCCCAGGG
         CCTTCGCCTGGGATCTGCACAGCTTTATAAGCAGCCCAGGGTGAGAGATGGGCCTGTCTG
         TGGTTCCCAGAGACCACGGCAGGAAATTCTCTGTCACCATCGGTGCATGGGCAGGGGCCA
         [G,A]
         AGAACCGGTGGCACAAGGTGTCCCTGGCTCTCTGCTCAACAAACAGCGAGTGCCCAGTGA
         CTGCGACCAGGCCCCGCTCTTGGGATGAGGACAACCGTCTGGGAACGTCCACGCACCCTT
         TATGAGACACAGCACGCGCCAGCACCGCAGTCACACACCGGGGGCTCGGGTCAGCCTCAT
         AGCTGCCCGGCCTTGAGTGCTGGGCCTGCGTCTGTGAGCAGCGCCCACCTGGGTGGCGGT
         GGTGTGTGCTTCACTTCCACACACTCCCGTGCATGCTCCCCGGCCTTCTGGGGTGGCTCG
```

Chromosome mapping:
Chromosome 10

FIGURE 3K